

Better government for Europe?

The Maastricht Treaty on the constitution of Europe from now on will probably be ratified in the coming months, but will not give Europe the better government it needs.

WHEN will Europe have a government worthy of the name? As the twelve members of the European Communities (EC) were embarking two weeks ago on the ratification of last December's Maastricht Treaty, which among other things will rename the EC the European Union, their agriculture ministers were producing like a rabbit from a hat a radical reform of the Common Agricultural Policy (CAP). The principle is that Europe's excessive production of cereals, meat and milk will be contained by paying farmers not to produce them. The device may be sufficient to rescue from failure the five-year negotiations on the General Agreement on Tariffs and Trade (GATT), but it is otherwise nothing but an ill-conceived compromise between the interests of national governments (not all of which are satisfied by it). There is no sense in which the people of Europe have been consulted on this expensive transfer of the cost of European agriculture from the pockets of consumers to those of taxpayers. That is not merely a shame, but a threat to the stability of the reform. Measures that have not been discussed cannot compel consent or carry conviction.

The reform of the CAP is not the only major EC decision to have been sprung on Europe's citizens in this way, of course. Decision-making by fiat is the rule, not the exception, deriving from the constitution of the EC. The European Commission, the over-maligned bureaucracy in Brussels, drafts legislation, which acquires legal force only when adopted by what is called the Council of Ministers — the collective name for the miscellaneous committees of national representatives, usually ministers with special interests, who swing into Brussels for a day or so to cast their votes (on those matters subject to majority decision since 1986) or to exercise their vetoes. The reform of the CAP was wrought over three days (including one all-night session) by Europe's ministers of agriculture — people most of all conscious of the potential influence of farmers in their own domestic politics, but not necessarily people with a sense of the wider interests of the national communities they represent. Nobody should be surprised that the outcome is a pig's ear.

Privilege

That European farmers are privileged is unsurprising. The Treaty of Rome, drawn up (in the 1950s) when it seemed that agricultural communities would be permanently at a disadvantage compared with others, guaranteed that farmers would be protected to some degree from the expected winds of change. The principal mechanism for this purpose has

been price-rigging; the prices of European farm produce have been fixed annually at levels calculated to yield supportable farm incomes. Because prices have inevitably been greater than world market prices, imports of food have had to be kept out of Europe by tariffs (called levies) and European exports have been possible only because of the subsidies with which they have been accompanied. Most of the rest of the world rightly objects that these practices offend against the principles of free trade, and that the general effect of the CAP, which is to sustain European agricultural production at artificially high levels, distorts the natural division of labour in an increasingly skill-centred world. Counting direct spending on buying and storing surplus food and the indirect effects of food prices higher than would otherwise be necessary, the cost of Europe's CAP is huge — some US\$50,000 million this year.

Remedy

The remedy the EC is touting will not much reduce this horrendous bill. Farm prices will come down (by 20 per cent for cereals, for example), but farmers will be compensated directly for the incomes it is estimated they would otherwise have earned. Farmers (especially in France, Germany, Italy and Greece) are up in arms, fearing that compensation will be neither sufficient nor equitable. What should be worrying the rest of Europe is that the proposed reform imposes a substantial and supposedly permanent tax burden on Europe's economy without meeting the legitimate expectations of Europe's agricultural communities. Specifically, there is no assurance that the compensation will be equitably shared between farms of different kinds, or in different countries; indeed, what constitutes equity has hardly been discussed. Similarly, there is no vision of how a transition will eventually be made from the new rigged market in food to the supposed end-point, a transparent market. Nor has serious attention been given to the question of how agricultural communities might be liberated from their present dependence on the cultivation of the land. But the most serious defect of the pretended reform is that it may permanently prevent Central and Eastern Europe from selling food to the West.

These are all complicated issues that can be resolved satisfactorily only by careful study and discussion in public. How else is it possible to command the understanding of a wide and diverse public such as that of Europe? Yet the standard practice of the Council of Ministers is quite the



opposite. Compromises are arrived at by horse-trading at hasty negotiating sessions. Not even the minutes of these meetings are made public, so that only rumour makes it possible to tell who said what, and why. Even sensible decisions lack conviction because of the melodrama of their making. There is no good reason why issues affecting the whole lives of Europeans should be left to the specialized ministers who attend the council. The latest decision on the CAP reform, for example, plainly reflects the special interests of people generally regarded as the sponsors of their domestic farming lobbies; by what right do these people legislate for the future pattern of Europe's economy, not to mention the continued growth of world trade?

Much the same has happened (and will continue to happen) in the EC's management of its research programme, on which it spends more than \$1,000 million a year. The underlying objectives have been from the outset of the successive 'framework' programmes a miscellany of ill-matched wishes: that innovative technology should prosper, that collaboration between companies and academic institutions should span European frontiers and that the mobility within Europe of technical people should be greater. In themselves the wishes are admirable, but it has never been seriously considered whether the framework programmes would meet those objectives. The practice of the past ten years is not encouraging.

It is a great misfortune that the constitution of the EC prevents it from reaching more intelligent decisions more intelligently. In principle, of course, more vigorous advance discussion of the European Commission's proposals by the European Parliament would help; in practice, the parliament has not yet won its intellectual and political spurs. For the time being, national parliaments are better equipped to scrutinize the plans the Commission is forever bringing forward. In reality, they behave as if they had neither the inclination nor the time to tackle any but marginal aspects of these problems. (Nevertheless, the analysis two years ago by the British House of Lords of the diversion of EC funds away from the intended purpose was a trenchant comment on an important problem.) But, given the Maastricht Treaty and the certainty that national policies will increasingly be framed in a European context, it can be held that they have no more important task. □

The decade of the brain

Arguments that research support for neuroscience should quadruple miss the point and are not convincing.

ACCORDING to a report produced for the National Foundation for Brain Research in Washington, DC, disorders of the brain (including neurological and psychiatric diseases) cost the United States \$401,000 million a year. The economic analysis in *The Cost of Disorders of the Brain*, prepared by the firm Lewin-ICF, breaks that figure down as follows: neurological disorders \$104,000 million, psychiatric disease \$136,000 million, alcohol abuse \$90,000 million and

drug abuse \$71,000 million. These substantial sums include the direct costs of medical care as well as indirect costs of the diseases, such as lost wages, and add up to 7.3 per cent of the US gross domestic product.

Armed with these impressive data, researchers affiliated with the brain foundation have initiated a public relations and lobbying effort to increase federal support for diseases of the brain, which at present come to a total of \$1,800 million. The foundation makes the argument that an expenditure of \$401,000 million justifies an outlay of research dollars roughly four times the present figure. An investment of \$8,000 million, he suggests, is entirely reasonable, especially because the savings in dollars that would result from a cure for Alzheimer's disease or for schizophrenia would more than repay the government's contribution to research. It is an appealing argument. Unfortunately, it is quite wrong-headed.

In the United States, where the government has a penchant for assigning various causes to a special week of their own, the entire 1990s are officially the decade of the brain. Fair enough. Advances in neuroscience during the past few years certainly validate a concerted effort to push back the research frontiers even further. New data suggesting that neurons, once thought incapable of regeneration under any circumstances, can repair or regenerate themselves, provide but one example of the excitement in the laboratory.

However, there is as yet no certain link between significant progress in understanding neuronal function and its application to any human disease. Would that it were otherwise. Thus, the promise of a cure for Alzheimer's, like the promise two decades ago that cancer could be defeated by declaring war on cancer from the White House, must be regarded as optimistic at best. In all likelihood it is a cruel overstatement.

The figures for the economic costs of brain and neurological diseases are themselves somewhat misleading. Unhappy though it may be, these diseases, which cost patients, families, insurance companies and governments \$401,000 million a year, also provide jobs for hospital workers, home care givers, construction companies that build nursing homes and so on. Given the profound anguish a disease such as Alzheimer's causes patients and families, it sounds cruel to suggest that there are two sides to the economic picture, but that is the case.

Scientists have long tried to cast their pleas for research funding in terms of dollars lost to disease. The brain foundation did not originate the notion. But it is an approach to be avoided. There are more than enough reasons to ask federal agencies to contribute large sums to brain research (first among them the fact that the science right now is ripe for development) but lowering the gross domestic product should not be listed among them. □

Correction

In a comment on the US Environmental Protection Agency's (EPA) position on the safety of recombinant DNA (see *Nature* 357, 179; 1992), Dennis D. Foote of the University of California at Riverside is the researcher who should have been credited with calling EPA to task for putting politics ahead of science.

Sandia looks for a company that will put research before profits

Washington. "Wanted: Manager for a US Department of Energy (DOE) national laboratory with dwindling Cold-War-related mission. Must be willing to restore tarnished image if chemical/radioactive waste problems are discovered. Compensation possible, but sense of patriotic duty preferred."

Later this year, DOE will issue an advertisement for Sandia National Laboratory that will look nothing like this, except to those who can read between the lines. American Telephone & Telegraph (AT&T), which has been running the Albuquerque, New Mexico, laboratory on a non-profit basis since the mid-1940s, announced last month that it would be pulling out next year (see *Nature* 357, 104; 1992). AT&T had accepted the responsibility after President Harry Truman made a special appeal during the Second World War to help propel the United States into the nuclear age. But when DOE goes looking for a replacement, it may find that patriotism is no longer enough.

Realizing that running a national weapons laboratory is often a thankless task, DOE officials say that they may allow AT&T's successor to run Sandia at a profit. If a university or university consortium decides to take over management of the laboratory (as in the case of Lawrence Livermore and Los Alamos national laboratories, which are both run by the University of California), DOE hopes to maintain the non-profit arrangement. But if a private company is the most attractive candidate, DOE is prepared to offer a deal that provides financial incentives beyond costs.

That possibility is already causing tension between the laboratory and DOE headquarters. Sandia officials are concerned that a search for profits may skew the delicate balance struck by allowing researchers as much freedom as possible within the confines of national priorities. Some incentive arrangements, which reward contractors for working closely with laboratory officials, prize cooperation above the quality of the research. And Sandia researchers are concerned that such an arrangement could compromise their independence and research objectivity.

"When we work with DOE now, they don't ever have to think about the possibility that we're being influenced by somebody's profit motive", says Lee Bray, Sandia's executive vice-president during the transition to a new contractor.

Thomas Ambrose, an official at the University of California who is renegotiating the management contracts for Lawrence Livermore and Los Alamos, says that incentive fees have their place, but not in manag-

ing science. One place where Ambrose thinks that a profit is appropriate, he says, is DOE's Oak Ridge National Laboratory, which produces nuclear materials and is run by Martin Marietta Corporation for a substantial fee. Research and engineering-orientated Sandia, however, is another matter.

"If you get an incentive fee for managing



**Sandia
National
Laboratories**

a production operation, that's not so bad", says Ambrose. "But if you start managing research and development on a fee basis, you can get into a conflict" between scientists and managers, he says.

Yet profits may be an essential element in any effort to entice a company to manage a weapons research laboratory. "It's a tough question", says Everett Beckner, assistant deputy secretary for defence programmes at DOE headquarters. "Clearly [this issue] has an overriding impact on the way the lab is regarded and managed."

Although some argue that working side

by side with laboratory researchers would give a managing company insight into technology that an outside company might not recognize, a contractor is allowed no special rights to laboratory technology. Sceptics note that of the 30 Cooperative Research and Development Agreements that Sandia has signed with private companies, none has been with AT&T.

Nevertheless, Sandia officials are optimistic that another company will find the laboratory's technology to its liking. AT&T left because its corporate direction no longer meshed with the priorities of the laboratory, they point out, not because there was anything wrong with Sandia's research.

But Beckner gives little weight to the idea that a manager would benefit much from an inside track on laboratory technology. "If a firm came into this hoping to drag a lot of ideas out of the lab and establish a nice business, it'd be fooling itself", he says.

Although DOE will not solicit bids until the autumn, several groups are said already to be thinking about the challenge. One is a new university consortium being organised by the University of New Mexico — just the sort of group (if it can convince DOE that it could adequately manage a \$1,000 million national laboratory) that would be expected to emphasize scientific freedom over seamless bureaucracy.

Christopher Anderson

Germans try to ease overcrowding

Munich. German university students in science and mathematics are facing crowded classrooms and waiting lists for popular courses as they cope with a system creaking under the strain of rising enrolments.

The problem stems from a policy that gives any student who passes the high school graduation examination the right to higher education. Although no one wants to abandon this democratic ideal, the crisis is forcing politicians to consider changes in how the system should be funded.

Higher education is financed by the regional governments of Germany's 16 states. They, in turn, depend on the federal government for their annual budgets. This situation has in recent years led to much rhetoric but little action.

That could be changing. The cultural and finance ministers of the regional governments, while repeating their desire to retain control of education funding in the long term, have asked the federal government for

DM3,000 million (US\$2,000 million) a year for new buildings and research facilities, more teachers and larger grants to students.

More than 30,000 new teachers are needed to restore student/staff ratios to the levels that existed in 1977, says Hans-Uwe Erichsen, president of the Committee of University Principals. At Ludwig Maximilian University in Munich, for example, student/staff ratios have risen from 7.7 to 13.8 in physics, from 11.7 to 18.9 in mathematics and from 6.8 to 16.6 in biology during the past 15 years.

Erichsen wants higher education officials to meet representatives of regional and federal governments to discuss the problem. The German chancellor, Helmut Kohl, has agreed in principle but has warned that it will be hard to decide who should pay more. Another group, the Committee of Regional Presidents, will discuss the issue this summer at a special policy meeting.

Allison Abbott

Science panel's report could end whaling ban

London & Tokyo. Next week the scientific committee of the International Whaling Commission (IWC) is expected to complete work on a document that makes the case for the lifting of a ten-year moratorium on commercial whaling. But a new French proposal to turn a portion of the southern ocean into a whale sanctuary may scuttle any attempt to restart commercial whaling.

The IWC, which meets later in the month in Glasgow, is bound to regulate the conservation and utilization of whale stocks on the basis of scientific findings. The moratorium was put in place in 1982 so that the perceived inadequacy of the science could be set right. Although there were loopholes that allowed some whaling to continue under the provision of research, the numbers caught were relatively small and bodies opposed to whaling felt that the scientific task was so complex that the moratorium could be extended indefinitely. This situation persisted past the IWC's original deadline of 1990.

However, it became clear last year that the scientific committee was close to completing its work on what is called a Revised Management Procedure (RMP). That is when things started to heat up.

In March, the scientific committee met in Copenhagen to iron out most of the existing problems at the core of the RMP, and to answer the IWC's outstanding queries. The management procedure sets out what data need to be collected and how they must be analysed in order to establish how many whales can be caught without threatening their survival.

Although there are different implementation rules for different stocks and extra procedures for multi-stock situations that still need to be resolved, the RMP is ready to be applied to the now well-documented stocks of Minke whales in the North Atlantic and the seas of the Southern Hemisphere. It is within these populations that nations such as Japan, Iceland and Norway want to resume whaling.

Few doubt that the RMP will be presented to the IWC meeting, which begins on 29 June. In recent years the scientific committee has drawn so much attention from both sides of the whaling debate that it is perhaps more balanced than the IWC itself.

What will happen at the main IWC meeting is less predictable, however. Despite the precept that the decisions should be based on science, the determining factors are likely to be political. The countries whose objections to whaling are essentially ethical have little room to manoeuvre in terms of extending the moratorium. There are still matters such as inspection and observation proce-

dures to be finalized, which could prevent commercial whaling for another couple of years, but attention is likely to switch to ways of limiting the extent and economic viability of the resumed commercial whaling.

Chief among these moves is a proposal from the French government that the waters south of 40°S should be indefinitely designated a whale sanctuary. This is presented as a scientifically motivated plan, as indeed it must be because the IWC does not permit sanctuaries to be established merely to invalidate catch limits.

There is a great deal of scientific background to the proposed sanctuary, the essential element being that it will contribute to the "rehabilitation of the Antarctic marine ecosystem". It would reinforce and complement other conservation measures, in particular the protection of all baleen whales in the Southern Hemisphere and the sperm whales in their feeding grounds.

Fuzuko Nagasaki, executive director of the Cetacean Research Institute in Tokyo and a member of the Japanese delegation, believes that the sanctuary is a "very political" proposal and that it has "no scientific basis". "No whaling, not even scientific whaling, could be carried out", says Nagasaki.

The IWC's rules of sanctuary also exclude the killing of whales for scientific purposes — a much debated loophole that pressure groups argue has allowed commercial whaling elsewhere under the guise of science.

Nagasaki says that the Japanese government is trying to stop the French proposal by lobbying the IWC member nations before the meeting. But if there is a vote, he

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Humpback whale off Newfoundland, Canada

suspects that the French could obtain the 75-per-cent majority needed to carry the proposal. If that happens, Nagasaki says, Japanese legislators are likely to call for Japan to leave the IWC.

It is possible that a strong research procedure built into the sanctuary plan could help carry it through the IWC, although at present the French proposal contains only a suggestion that the scientific committee be asked to look into the matter. The outcome could hinge on attracting the support of Australia, New Zealand and the South American countries whose territorial waters impinge on the proposed sanctuary.

Japan has lost an important ally in Iceland, which has announced its intention to leave the IWC at the end of June and will not be voting in the meeting this year. Iceland may have pushed itself into a corner by leaving the IWC: to resume whaling and remain within the United Nation's Law of the Sea, it has to co-operate with the relevant international organizations — until recently only the IWC. However, in April it became one of the founding members of the North Atlantic Marine Mammal Commission along with Norway, Greenland and the Faeroe Islands.

The new commission has said it intends to manage only small cetaceans such as small whales and dolphins. But the potential exists to set catch limits on larger whales.

Ian Mundell & David Swinbanks

France welcomes British physicists

London. British nuclear structure physicists, soon to be deprived of an accelerator to work with in their own country, can now transfer their affections to France. An agreement signed by Sir Mark Richmond, chairman of the Science and Engineering Research Council (SERC), and Claude Detraz, director of the Institut National de Physique Nucleaire et de Physique des Particules, provides a framework for collaboration in nuclear structure physics.

In general, the agreement gives French scientists access to British funds and equipment and British scientists a place to work. Since the closure next year of the Nuclear Structure Facility (NSF) at Daresbury was

announced in 1991 as a cost-cutting measure, international collaboration has been seen as the only salvation. The SERC hopes that the agreement will ease that process.

Eurogram, an ultra-sensitive gamma-ray detector built at the NSF, will be used for the first project under the new agreement. Part of an existing collaboration, it was already due to be transferred in 1993 to the Vivitron tandem Van de Graaff generator being built in Strasbourg. The second project is likely to involve Charissa, an NSF-based array for detecting charged particles. It may also travel to France, although other sites in Europe and Australia are still being considered.

Ian Mundell

Italian space scientists win round in funding

Munich. The Italian government last week came down on the side of space scientists in their battle to recover millions of dollars now going to commercial space projects.

Antonio Roberti, Italian research minister, has told the Italian Space Agency (ASI) that it must adhere to a law that requires the agency to spend 15 per cent of its budget on research. The president of ASI, Luciano Guerriero, has been deducting Italy's contribution to the European Space Agency (ESA) — of which only about one-seventh goes for research — and other costs before calculating the share that researchers are entitled to receive.

The problem, which has been building up since ASI was created in 1988, came to a head last year when the agency's science advisory committee blocked the agency from spending any of its 1991 budget in protest at ASI's policy. Its actions effectively paralysed many international research collaborations (see *Nature* 356, 647; 1992). Roberti also reprimanded Guerriero for funding some research projects without first consulting the advisory committee, a further contravention of the regulations.

The minister's intervention vindicates the committee's chairman, Remo Ruffini of the University of Rome. Ruffini had been criticized by scientists and by members of his own committee for exercising his right to refuse to approve any award in order to increase the amount of money going to research. Committee members now say that they are united behind Ruffini and have

dropped their demand that he resign.

Roberti, in his directive to Guerriero, supported Ruffini's tactics. He said he was "very sensitive to the unhappiness of the scientific community" and he demanded that ASI resolve the issue immediately. Roberti also said that ASI's decision to fund some projects on its own was unconstitutional. The agency's support for SAX, an expensive X-ray astronomy telescope, despite the advisory committee's belief that the science is out of date, had contributed to the breakdown in relations between the two bodies.

Ruffini will now allow ASI to spend all of its available research money. That amount, some 46,000 million lira, falls well short of the 68,000 million lira that researchers had expected to receive in 1991 and the request for 74,000 million lira that the committee made last autumn in the midst of its battle with ASI.

But the dispute is far from over. Ruffini wants the government to restore all of the funds withheld by ASI from researchers through the years. Guerriero says that the question of retrospective funding is being reviewed.

Guerriero also denies that his interpretation of the law penalized scientists, and says that the difference, if it exists at all, is trivial. "International research programmes paid for from our ESA contribution will still have to be deducted from the 15 per cent", he says.

Ruffini, in turn, accuses Guerriero of taking a cavalier attitude towards govern-

ment funds, and insists that the additional money due to science is substantial. "It will probably be around 40 billion lira", he says, an amount Guerriero disputes.

The debate over money obscures what some scientists believe is a more serious problem facing ASI — its system of awarding grants. Giovanni Bignami of the University of Cassino, whose work has been delayed because of the infighting, says that fundamental changes are needed. "Now that the ministry has taken a firm stand on how much money [for research] should be given, it must also find a means to divide the money in a professional and mature way", he says.

Bignami would like to see ASI use the same type of independent peer-review panels that are common in other space agencies. At present, all work funded by ASI is selected by the advisory committee, which includes researchers in a variety of applied areas such as telecommunications and robotics. The more fundamental scientists feel that their views are being ignored, and they fear that those with commercial interests may be exerting too great an influence on the selection process.

Robertis said that the ministry will not consider ASI's provisional budget for 1992 — submitted last month, seven months late — until ASI has solved its research funding problems. He has given ASI two months to put its house in order. In the meantime, the committee is developing its recommendations on the assumption of a significant increase in its funding. **Alison Abbott**

Europe decides manned space costs too much for now

Munich. Plans for Europe's manned space programme have been shelved, possibly forever, as the project becomes too costly for member states of the European Space Agency (ESA) to afford.

Jean-Marie Luton, director-general of ESA, announced last week that his agency will recommend a postponement until at least 2005 of plans for the space plane Hermes and the man-tended free flyer that is part of the space station Columbus. Instead, the technology developed so far will be displayed on an unmanned demonstrator to be launched in 1998 with the hope that it will attract investors from outside Europe. The missions were originally scheduled to be ready in 1998.

In the meantime, the ESA will concentrate on its Earth observation projects, in particular those that monitor the environment. Those projects will benefit from the US\$700 million to be saved by postponing Hermes and the free flyer.

The recommendation is expected to be

adopted by ESA delegates meeting later this month in Paris. It would then be officially approved at a ministerial meeting in November in Spain.

The announcement came as no surprise to the space research community. Rumours had been circulating for months that an independent European manned space programme was no longer feasible in the current economic and political environment. The cost of Hermes has risen by 40 per cent in five years, to a current estimate of US\$7,500 million.

Last November, ministers from ESA member countries voted to delay the Hermes and Columbus projects for one year to give them time to reconsider their rising costs and to search for new partners. That action followed a decision by the German government to reduce its funding in real terms.

The immediate problem arose after the German space agency, DARA, was told that its annual budget would be increased by no more than 2.5 per cent in the next few years,

well below the 10 per cent rate that is needed to meet the country's commitment to European space programmes that Germany made in 1987. That left ESA with a shortfall of about US\$650 million, a gap that other member states were unwilling to fill.

But the president of DARA, Wolfgang Wild, says that economic considerations were not the only reason for delaying the projects. "The political changes in eastern Europe have opened up new possibilities for cooperation that we must investigate", he says.

ESA members have spent the past few months trying to forge links with other countries, specifically Russia. ESA hopes that tapping into Russia's expertise in manned and unmanned space exploration will lower the cost of its own programme.

"The man-tended free flyer is not so important now if Europe can have access to the Russian space station Mir", says Wild. Still, overtures to the newly formed Russian space agency have yet to produce any tangible results. **Alison Abbott**

Genetically engineered foods get green light

Washington. Last week the US Food and Drug Administration (FDA) embraced the principle that the government should regulate the products of genetic engineering, not the process by which they are created. Its proposed guidelines for the regulation of new varieties of foods, which had been expected for months, represent an unofficial endorsement of the idea that government agencies charged with preserving public health do not need special rules to oversee genetically engineered organisms.

The argument that such special treatment is unnecessary has been made by scientists and the biotechnology industry in several reports over the past few years*. It has not convinced environmental groups, however, which argue that the safety of the nation's food supply is being compromised by an eagerness to help companies bring new products to market.

The proposed new guidelines state that foods developed using genetic engineering pose no new or special safety risks to the consumer and should be subject to the same standards of regulation as are applied to all other foods. In a joint announcement by Vice President Dan Quayle, chairman of the president's Council on Competitiveness, and the Department of Health and Human Services (HHS), FDA's parent agency, HHS Secretary Louis Sullivan said these new technologies will provide foods "that are tastier, more varied, more wholesome and that can be produced more efficiently".

The guidelines are consistent with the line taken by the competitiveness council, which has pushed for the elimination of unnecessary regulatory burdens on the industry. They are also in keeping with similar measures introduced earlier this year to streamline the approval process for biotechnology drugs.

The guidelines put the onus on industry

officials to decide whether a new genetically engineered plant variety needs pre-market approval by FDA. FDA is the primary government agency with responsibility for ensuring safety of the nation's food supply, with the exception of meat and poultry.

Several consumer protection groups are alarmed by the notion of industry policing itself. Jane Rissler, biotechnology specialist with the National Wildlife Federation, called the policy "outrageous". "It is not protective to have a voluntary programme for an industry that is just so competitive and trying so hard to get the first products to market", she says.

At the same time, the long-awaited guidelines are welcomed by the agricultural biotechnology industry, which lags far behind other sectors of the industry and has yet to bring a genetically engineered food to market in the United States. According to the industry's trade group, the Industrial Biotechnology Association, more than 50 crops produced through genetic engineering have been tested.

Under the new guidelines, foods developed through genetic engineering, including fruits, vegetables and grains, will be regulated within the existing framework of the federal Food, Drug and Cosmetic Act. The level of oversight will be based on the characteristics of the food and its intended use, rather than the method by which it was produced.

FDA expects that many of the genetically altered foods being introduced (as with most new plant varieties produced by more traditional breeding methods) will not require pre-market approval by the agency. However, pre-market approval and labelling will be required if the use of genetic engineering is shown to increase the concentration of a naturally occurring toxicant in the plant, to introduce an allergen (protein that can trig-

ger an allergic response in some people) that is not commonly found in the plant or to alter the nutritional composition of the plant. Under the new guidelines, industry executives will need to pick their way through a series of "decision trees" to decide whether to consult FDA when introducing a new genetically engineered plant variety.

Predictably, the announcement was criticized by consumer protection groups, which see it as another example of the administration meddling in the affairs of the FDA. They believe that it does more to enrich the agricultural biotechnology industry than to protect consumers.

The public interest community "is getting beaten about the head", says Rebecca Goldberg of the Environmental Defense Fund. "When industry's interests differ from that of the public, the public does not win", she says.

On the day the new guidelines were announced, Jeremy Rifkin, president of the Foundation of Economic Trends and a persistent critic of biotechnology, filed a legal petition with FDA demanding that genetically altered foods be labelled as such. Rifkin wants foods containing new genetic material to be regulated in the same way that FDA regulates new food additives.

A frequent target of Rifkin's attacks, Henry Miller of the FDA's Office of Biotechnology, defends the proposed guidelines as a logical progression. He says that foods have been engineered in ways that transcend the normal barriers of reproduction for the past two decades, with genes being moved from one species to another and even from one genus to another.

One of the first genetically engineered foods to reach the grocery stores is likely to be the new Flavr Savr tomato from Calgene, Inc. of Davis, California. Researchers at Calgene have isolated from tomatoes the gene that encodes the polygalacturonase enzyme, which causes fruit to soften. Reinserting the gene into tomatoes in a reverse or "antisense" orientation blocks as much as 99 per cent of the enzyme's production.

Interested parties have 90 days to submit comments to the FDA. But barring a groundswell of public opinion to the contrary, the proposed guidelines are likely to stand.

Diane Gershon

Indian missile test fizzles out

New Delhi. The second test of India's intermediate range ballistic missile, which the United States tried hard to prevent, ended in failure last week. The launch of the Agni was normal, according to the Indian Defence Research and Development Organisation, which developed the missile, but "its mission objectives were not fulfilled due to premature ignition and separation of the second stage". A third test, planned for next month, is expected to proceed on schedule.

India's decision to go ahead with pre-production testing of Agni is thought to be a sign to the world that the country's military missile programme, unlike its space activities, is immune from threats of em-

bargo by countries that subscribe to the Missile Technology Control Regime (see *Nature* 357, 271; 1992). A few days before the launch, the country's defence minister, Sharad Pawar, said that India will not abandon the field even if the country is shunned by the industrialized world.

Agni, which means fire in Sanskrit, flew first in 1989. It has a range of 2,500 km and can carry a one-tonne payload. Its first stage uses the same solid-fuel booster as the SLV-3 rocket being developed by the Indian space agency; the second stage is a twin-engine, liquid-fuel propulsion system similar to what it hopes to use to launch commercial space satellites.

K. S. Jayaraman

* *Field Testing Genetically Modified Organisms: Framework for Decisions*, Report by the US National Academy of Sciences/ National Research Council, (Washington, DC, 1989); *Strategies for Assessing the Safety of Foods Produced by Biotechnology*, Report of a joint Food and Agricultural Organization (FAO)/World Health Organisation (WHO) consultation (Geneva, 1991); *Exercise of Federal Oversight Within Scope of Statutory Authority: Planned Introductions of Biotechnology Products into the Environment*, Document by the Office of Science and Technology Policy (OSTP), (Washington, DC, 1992).

European Environment Agency still in limbo

London. When Carlo Ripa di Meana, the environmental commissioner of the European Communities (EC), announced that he would not be going to the current Earth Summit in Rio de Janeiro because environmental policies should be "based on binding obligations and precise undertaking, not on words", he might as well have been talking about events closer to home. Two years after the European Environment Agency (EEA) was created, its only outcome has been angry words, specifically a continuing political squabble about where it should be located and what it should be doing.

A report published last week by the British government's Advisory Council on Science and Technology (ACOST)* deplores a situation in which negotiations about the site of the European Parliament have stood in the way of the EEA starting on its job of collecting, collating and issuing environmental data. To make matters worse, the EC's existing programme for gathering environmental data was brought to a halt when the staff responsible were transferred to the task force that was supposed to set up the agency.

Fortunately, that task force, led by the EC's Philip Bourdeau, has managed not only to lay the foundations for the EEA but also to continued the work of CORINE (the

information system on the state of the environment in Europe). Although the task force is thus carrying out part of the agency's intended function, one important element of what the EEA could do is still missing.

"The EEA has a primary responsibility to be the repository of environmental data for Europe", says Robert May, chairman of the ACOST working party on environmental research programmes. "But rather than just taking the data it is offered, it should also be thinking about what data are needed."

The EEA was never intended to fund research. But its creators hope that it will eventually have the ability to direct environmental monitoring and research programmes to fill gaps in the European dataset. By the same token, it could reduce the duplication of work that arises where nationally coordinated programmes meet at borders and set standards to ensure the compatibility of data from different sources.

Several European countries, notably the Netherlands, recognize the need for comprehensive registers of monitoring within their borders and have taken steps towards achieving that aim. However, without the EEA or a similar agency, the result is a patchwork of unrelated information.

Although individual nations have differing perceptions of the division between

research and routine monitoring, data from both activities would feed into the EEA, and everyone would benefit from the baseline data that the EEA would provide. So although arguments may arise over who should fund which part of the work and whether the EC is spending its money wisely, there is every reason for the EEA to proceed.

The ACOST report also recommends, in addition to establishing a register of environmental monitoring activities, that Britain increase spending on local and regional environmental research and shift the emphasis from physical processes to the impact of environmental change. In particular, it calls for more research on understanding how physical and chemical changes in the environment affect and are affected by biological processes.

Bourdeau is hopeful that the European heads of state will address some of these issues at their summit meeting later this month. But it should be remembered that Britain, which will be taking on the presidency of the EC at that summit, just two months ago abandoned plans to establish its own environmental agency. **Ian Mundell**

* *Environmental Research Programmes*, by the Advisory Council on Science and Technology (HMSO, PO Box 276, London SW8 5DT, or call 44-71-873-9090).

Eureka welcomes first East European member

Paris. Hungary has become the first East European member of Eureka, the high-technology programme of the European Communities (EC). Filippo Pandolfi, vice president and head of research for the EC, announced the decision two weeks ago at a Eureka meeting in Tampere, Finland, that coincided with the rotation of the programme's presidency from Finland to France.

Last June, Eureka announced that it would extend membership to countries in Eastern Europe. Since then, Eureka has established contacts with Hungary, Poland, the Czech and Slovak Federal Republic, Romania, Estonia, Lithuania, Albania, Slovenia and the countries of the Commonwealth of Independent States (CIS). Eureka officials say that they expect Poland and the Czech and Slovak Federal Republic to become full members of Eureka within the next 12 months, although neither country has yet applied formally.

Although non-member countries can take part in Eureka projects, the entrance criteria are more restrictive than for members. The new partner must take a major role in the project, and research must be carried out mainly in the member countries. Adminis-

trative procedures are also more cumbersome. In 1991, some 17 Eureka participants in 10 projects were from East Europe.

Hungary, the first East European country to have officially applied for membership, has been involved with Eureka for longer than any of its neighbours. It is also the first country to meet such entrance requirements as having "an active market economic policy" and national research institutes and companies that have been involved with Eureka.

Hungary is expected to participate more fully in existing Eureka projects now that it is a full member; it already has limited involvement in several Eureka projects under the rules for outside participation. In the longer term, access to the Eureka network of Western companies, institutes and research projects should help Hungary to build research networks, break into new markets and commercialize its technologies. Eureka membership also improves the chances of Hungarian companies obtaining cash from their government for research. In turn, Hungary will begin to contribute to Eureka, paying as much as 35 per cent of any project in which it participates.

This enlargement of Eureka illustrates

the continuing expansion of the frontiers of European research programmes. Last year, the European Commission opened up all of its Environmental Research Programme to the countries of eastern Europe, and admitted Hungary, Poland and the Czech and Slovak Federal Republic to the EC's COST (Co-operation in Science and Technology) programme.

Some issues relating to participation remain unresolved, however, a result of differing research philosophies. Eureka, with a tiny secretariat in Brussels, helps companies seeking government funding to develop their ideas for joint projects. Most of the projects in the EC's Framework programme, on the other hand, are designed by the commission and then carried out by consortia.

Industry would like the EC to put more money into Eureka, but the EC wants Eureka to remain independent. Despite those differences Eureka and the EC agree that the EC should co-fund Eureka projects where there is a link with the EC's precompetitive programmes, and that researchers in other EC projects should be encouraged to develop this research further within Eureka to foster its commercial applications.

David Bakewell

Is peer review unbiased?

Sir — The thrust of your leading article "Conflict of interest revisited" (*Nature* **355**, 751; 1992) is based on the false premise that academic researchers not associated with the business world are unbiased expert participants in peer review panels of national agencies. Bias is a matter of degree, and whether conscious or unconscious it depends heavily on the goals, rewards and penalties that shape the framework of the expert's views. The academic researcher seeks the accolades of his professional colleagues, and depends on such recognition for promotion, tenure, professional awards and grants from government agencies and foundations. The politics of science reveals the extent of the manipulations by individual scientists to achieve these rewards. For most academics, these are more persuasive inducements than the money flow from business affiliations. The recent flurry of investigations into the ethics of scientists discloses the great value placed on these non-monetary goals.

Expertise in scientific specialties resides in those who work in depth in their fields, whether in academic institutions or industry. Such work almost always requires substantial financial support, either from government agencies and foundations or from industrial organizations. All these funding groups have goals that they assume will be furthered by the scientists they support, even though all avow that they seek only the truth. It is generally recognized in academic circles that government agencies are not likely to support those whose opinions might weaken the agencies' budget submissions to their government. Similarly, industry is very uncomfortable with research findings that question its public positions.

Does this pose a serious problem for advisory boards and peer review panels, as your leading article stated? Not if the myth of unbiased academic opinion is recognised. These boards and panels need the very best of expert knowledge, wherever it resides. The individual views and opinions that arise from each scientist's working relations are best disclosed and balanced by assuring a mix of experts. It must be recognized that every scientist, whether in an academic institution or industry, places maximum value on professional credibility. For this reason, in a mixed expert group a consensus on fact finding is usually arrived at without stress. Bias based differences usually occur in the area of policy implications, and it is here that the virtues of a mix of backgrounds becomes evident in disclosing the range of views of the various stakeholders. That is what

decision-makers need to know to balance national interests. So the problem your leading article poses would not exist if the self-serving image of academic objectivity was not so deeply embedded in academia and national institutes.

Chauncey Starr

*Electric Power Research Institute,
Palo Alto, California 94303, USA*

SLAC defended

SIR — I write in response to the two articles "Z-not" and "Zero-sum game" (*Nature* **356**, 99 & 97; 1992). The Stanford Linear Collider (SLC) had two major goals — to develop a new technology for electron-positron colliding beams that would allow the eventual construction of much higher energy machines at much lower costs than the old storage-ring technology, and to do some physics with this machine. The SLC has achieved its first goal, proving the feasibility of the technique to the point where large groups in Europe, Japan and Russia are now working with us to develop the technology for future linear colliders. Further experiments scheduled on the SLC are crucial to this international programme.

With regard to Z production, no one at SLAC to my knowledge has ever thought that the SLC could beat LEP in the raw number of Zs produced. Our niche was always going to be polarized beams. Although we have been late in reaching this goal, we agreed to a new set of milestones with the Department of Energy (DOE) in May 1991, and have so far exceeded all of them. We may fail, but haven't yet.

The other article in the same issue refers to a proposal that we have made to the DOE to build a B-Factor which the US high-energy physics community has said is extremely important to science. "Understanding the origin of CP violation is one of the central goals of particle physics," wrote the High Energy Physics Advisory Panel last year, one that "can be addressed in much more depth by the e^+e^- machines than by hadron colliders". There is also great interest elsewhere in such a facility, as evidenced by major groups in Europe and Canada that have expressed an interest in collaboration on experiments at the B-Factor. Unfortunately, there can be no funds available for such a project from the National Science Foundation until 1997 at the earliest, and no new DOE money for the foreseeable future.

In such circumstances, SLAC and its user community arrived at the proposal

to do the B-Factor out of the current budget by considering how to do the best possible physics for the funds available. SLAC is on target to meet its new goals, and we expect to be producing some unique physics next year.

Burton Richter

(Director)

*Stanford Linear Accelerator Center,
PO Box 4349,
Stanford, California 94309, USA*

Brownian motion

SIR — Cadée¹ disagreed with my argument that Robert Brown did not observe brownian motion², noting "... that his [Brown's] particles were too large (for example pollen grains). . .". The phrase in parentheses is Cadée's, which he then used to negate my contention². Brown reported vigorous motion with both 1/4,000–1/5,000-inch (6–5 μm) and 1/30,000-inch (1 μm) particles and with bismuth, manganese and even lead particles — clearly he did not observe proper brownian motion. Cadée's authority, Ford³, reported only a few qualitative experiments, compared to Jean Perrin's well-known, comprehensive, quantitative work⁴.

Cadée suggested that repetition of Brown's methodology with one of Brown's original microscopes, as Ford did, proves that Brown observed brownian motion. I doubt the interpretation, not the observation. Particle motion in Brown's methodology is too vigorous by orders of magnitude to be proper brownian motion. Following Ford, this simple demonstration illuminates the problem: a small drop of 10:1 diluted India ink is examined, uncovered, with a modern compound microscope, with cross-hairs in the eyepiece, at 100 $\times$. Vigorous motion of the particles will be seen. But if another drop is covered with a cover glass supported by two cover glasses, to form a cell 0.1 mm deep, virtually no motion will be observed. Pseudobrownian motion might describe the particle motion reported by Brown and Ford, whereas brownian motion is retained for the particle motion recorded by Perrin.

My serendipitous discovery regarding Brown's work arose from my 15-year examination of some of the foundations of physics. I am preparing a lengthy historical paper on Brown's work and brownian motion.

Daniel H. Deutsch

*141 Kenworthy Drive,
Pasadena, California 91105, USA*

1. Cadée, G. C. *Nature* **354**, 180 (1991).

2. Deutsch, D. H. *Bull. Am. Phys. Soc. Abstr.* **36**, 1374 (1991).

3. Ford, B. J. *Single Lens* 149–150 (Harper & Row, New York, 1985).

4. Perrin, J. *Brownian Movement and Molecular Reality* 1–93 (Taylor & Francis, London, 1910).

FBI's case for genetics

SIR — Your article "FBI gives in on genetics" (*Nature* 355, 663, 1992) is incorrect in its representation of the purpose of our recent collection of population frequency data from several European forensic laboratories.

Since the inception of the Federal Bureau of Investigation's DNA programme, our scientists have been engaged in a continuous assessment of the issues raised by DNA typing and its forensic application. The data that have been collected (some published in the scientific literature and some unpublished) are but further evidence of the reliability and validity of the FBI's estimates of the likelihood that a particular multilocus DNA profile will occur.

Whether using general population data for Caucasians, African Americans, Hispanics and Orientals, or for subgroups of these populations, it is generally accepted in the scientific community that an individual multilocus DNA profile is extremely rare and therefore highly informative for individual identification. Even so, the FBI has never assumed that there are no differences between subpopulations, but rather has taken a conservative approach that minimizes the effect of such differences as there may be between subpopulations.

The European and other population data we have been collecting are not new

to forensic scientists who attend the appropriate meetings. Our experience with such data forms the knowledge base on which is founded our confidence that subpopulations are not an impediment to the calculation of valid frequency estimates for DNA profiles.

The recent trip by FBI scientists to European and other laboratories was not undertaken in order to "re-examine those assumptions (of heterogeneity)", but was an attempt to accommodate those few critics of DNA profiling who suggest that the subpopulation issue is a source of concern.

Your article also suggests that incompatible laboratory standards in various countries could vitiate comparisons, but that statement does not acknowledge the international collaboration on DNA typing protocols that has been under way since 1988, notably by the Technical Working Group on DNA Analysis and Methods in the United States and the European DNA Program. This collaboration is but a further indication of the worldwide consensus which now exists on the methods and population sampling approaches in forensic DNA testing.

John H. Hicks

US Department of Justice,
Federal Bureau of Investigation,
Washington, DC 20535, USA

Ethical problems in the genes

SIR — The leading article "Genetics and the public interest" (*Nature* 356, 365; 1992) ends with the declaration: "the present spate of new knowledge in human genetics deserves the warmest welcome". Truly so, in terms of what it can tell us about the human condition and to help in understanding and alleviating human suffering. But not if some of the views you express are to become the basis for dealing with those who in the natural lottery of inheritance have drawn defective genes. What norm of fairness does your leader writer abide by that equates the self-inflicted risks associated with smoking with the hapless inheritance of aberrant genes? That your journal could make such a comment suggests to me that this issue of justice has not been adequately addressed by society in general or by bioethicists in particular.

If the natural lottery of inheritance is seen as "merely unfortunate" and not "unfair", there is no moral imperative for society to care for those with aberrant genes and their phenotypic expression. What guarantee, then, that "liberal societies should acknowledge an obliga-

tion to such individuals" as you suggest? I am concerned that in a climate of limited health care resources, such obligations will be readily overlooked. Englehardt in *The Foundations of Bioethics* (Oxford University Press, 1986, page 340) states: "the natural lottery creates inequalities and places individuals at disadvantage without creating a straightforward obligation on the part of others to aid those in need".

Perhaps this new genetic knowledge will become the driving force for social change. Our essential humanity will become in significant part the sum of our measurable imperfections. Then we can say let him (or her) who is without aberrant genes cast the first stone. Until then we need to proceed with circumspection and caution in our use of genetic information. The suggestion that the ethical problems of genetics are often insubstantial understates the problems in their social context.

David Turner

Haematology, Flinders Medical Centre,
Flinders University,
Bedford Park,
S.A. 5042, Australia

Interest in CNR

SIR — I was granted a fellowship by the European Communities to carry out experimental research in the field of molecular biology at a CNR (Consiglio Nazionale delle Ricerche) institute in Milan from 1991 to 1992.

Since January, my monthly salary, 1,938 ECU, and a bench fee of 5,000 ECU for the host institution have been sent to CNR headquarters in Rome quarterly, but the money has not yet reached its destination, the fellowship holder. I have been sustained partly by support from private companies, partly from the budget of the laboratory here in Milan. My monthly income has been highly unpredictable and I have not been paid in full.

This, apparently, is not an uncommon occurrence. CNR fellowship holders can wait several months before receiving any money at all and reimbursement of expenses for travels, conferences and so on may take longer than a year.

My question is: who is receiving the interest on this money, and when will somebody take an interest in changing this situation, which borders on theft?

Gertrud Lund

IBV, 15 Via Bassini,
20133 Milan, Italy

Not by numbers

SIR — It was a nice coincidence that Lord Kelvin's old chestnut "when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind" should appear on your Correspondence page (*Nature* 356, 653; 1992) next to a letter by Stephen Jay Gould (albeit on a different subject) who so elegantly refuted in *Wonderful Life* such a narrow view of science: "We will never be able to appreciate the full range and meaning of science until we shatter the stereotype of ordering by status and understand the different forms of historical explanation as activities equal in merit to anything done by physics and chemistry. . . . The answer to such questions as 'Why can humans reason?' lies as much (and as deeply) in the quirky pathways of contingent history as in the physiology of neurons."

J. R. Parker

Shell-Mex House,
Strand, London WC2R 0DX, UK

SIR — Science explains the maximum number of phenomena with the minimum number of hypotheses.

Brennig James

Cherry Orchard,
Marlow Common,
Bucks SL7 2QP, UK

What frontiers for Frontier?

Akiyoshi Wada

The Human Frontier Science Program is a bold attempt to support the next century's biology research. Its horizons should not be limited.

THE Human Frontier Science Program (Frontier), a Japanese initiative supporting international research on the brain and the molecular mechanisms of biological functions, has reached a critical juncture. After the first three-year phase, funded largely by Japan, the programme is now attracting financial support from many countries and its future direction is beginning to be questioned. As one of the principal architects of Frontier and as a former member of its council of scientists, I would like to explain the philosophy behind it and argue that it should maintain its basic principle of promoting pure life science by actively encouraging cooperation among many research fields in various disciplines.

Until the beginning of the Meiji restoration in 1853, Japan had isolated itself from the rest of the world for more than 200 years and the country did not participate in the development of science and technology happening elsewhere. Nevertheless, a strong education system had been maintained in Japan, and there was a reservoir of intellectuals able to take advantage of the new knowledge in science and technology that became available at the time of the Meiji restoration. British observers correctly evaluated the progressive engineering education system in Japan in those early days. Two articles in *Nature* (16, 44; 1877 and 71, 150; 1904) explained that in England technical education was conducted without scientific education to produce craftsmen, whereas in the rest of Europe, in particular in France, emphasis was placed on theoretical science without practical training. They then reported and praised the fact that for the first time an engineering education system with a well-balanced mixture of science and engineering had been established at the Imperial College of Engineering, the present Faculty of Engineering at the University of Tokyo.

I admit that Japan subsequently put too much emphasis on technology, particularly on areas with immediate applications. But at the same time I want to make it clear that about 120 years ago Japan established the basic concept that the fusion of science and technology, and balanced cooperation between them, are indispensable for the true development of both. This principle lay behind the establishment of Frontier,

which I think represents the Japanese genius at combining apparently different disciplines to create new fields of study.

The aim of Frontier is to help researchers in all fields to understand the true nature of life. In concrete terms, it provides research grants, long- and short-term fellowships and supports workshops, with a total annual budget now standing at about \$34 million (soon expected to rise to nearly \$40 million with contributions from the United States and other countries). In addition, it emphasizes the importance of supporting young researchers.

The purpose of the programme is to clarify the functions of living organisms by a 'molecular level' approach and by understanding brain functions. There are two 'priority fields': first, expression and transfer of genetic information, morphogenesis, molecular recognition and responses, and energy conversion; and second, perception and cognition, movement and behaviour, memory and learning and language, and thinking. Popular fields with their own societies and journals, such as cellular biology, immunology and developmental biology, were intentionally avoided. Through such an approach the intention was to encourage researchers in fields other than biological science to participate, as it was felt that knowledge can be acquired only when researchers are not confined to narrow frameworks.

Some participants and organizers of Frontier, unaware of this fundamental philosophy, have tried to narrow its priority fields to the above-mentioned conventional fields by claiming that 'molecular' fields are too broad and that there are too many research proposals to review. Such action is contrary to the basic philosophy of the programme. These priority fields were selected after careful consideration of how life-science research might develop without being confined to the framework of present-day research and should not be changed for trivial reasons, such as the difficulty of evaluating proposal applications.

I admit that there have been misunderstandings about the programme in the past. The fact that it was initiated by technologically advanced Japan and that its priority fields appeared to be oriented towards technology meant that some people suspected it was a strategy for Japan's technology development. But

now that an international peer-review system is functioning effectively, and the fairness of the system is well established, these doubts have been dispelled. Also, the intellectual property belongs to its creator(s), so Japan will not be the sole beneficiary.

I have said Frontier provides opportunities to examine the complex functions of life via an all-out effort of science and technology. Physics or engineering technologies (computers, robotics, electronics, materials science and so on) should provide powerful tools. Their contributions are not limited to hardware; the mode of thinking in fields such as mathematics and physics, information-science technology and engineering should stimulate the creation of new concepts in biological research. It is true that methods alone will not produce good research, but it is also a characteristic of modern science that highly sophisticated tools are indispensable for its evolution.

I regard biology as having entered a new phase in which it can develop with highly sophisticated methodologies, and believe that its progress can be quickly accelerated by a balanced perspective which views science and technology as one entity. I am convinced that Frontier is a unique project in which ideas and methodologies, science and technology, and physical and biological sciences, can be fused to promote life-science research. What is more, it was initiated by Japan, a country that had not previously been a significant international leader in basic science. Therefore, I believe that many researchers in various scientific and technological fields should join Frontier, and I firmly believe that it should continue to be open to and welcome their participation. Finally, as someone who has been involved from its inception, I rejoice in the fact that Frontier is developing into a truly international project, thanks to the people who have been connected with it — in particular those in the Secretariat in Strasbourg — and to the recent financial contributions from many countries. I ask researchers of the world to foster the healthy growth of this new-born baby. □

Akiyoshi Wada is director of the Sagami Chemical Research Centre, Nishi-Ohnuma 4-4-1, Sagami-hara Kanagawa, 229 Japan.

Russian summer hopes dwarf doubts

Russian science seems to have survived the winter and the first phase of price liberalization, but there is a long way to go before it can compete on equal terms.

Moscow. The difference between summer and winter here is not just the external temperature, but people's frame of mind. Especially towards the end of the long winter, spirits are at nadir. The return of a little natural warmth has each year liberated the natural optimism of those who live in this still-isolated part of the world. That, no doubt, is part of the explanation of the relative cheerfulness with which many research people now regard the future.

But there is also some substance in the change of mood since, say, last December. Seemingly insuperable problems have often been overcome by ingenuity and a little luck. The exodus of able and mostly senior people, for example, has not brought all the catastrophic consequences predicted.

The Moscow Institute of Molecular Genetics illustrates what has happened. Out of a total of 200 scientists and 150 students, no fewer than 40 people have left, of whom 36 previously occupied senior research positions. Many of the departed have finished up in US universities. Of necessity, these departures often decapitated successful research groups, leaving students and others leaderless at short notice or none.

Opinions are sharply divided on the manner in which some have left. Sometimes people have gone to "visit relatives" in the United States, and have allowed their intention not to return to be inferred only from the length of their absence. Some vigorously condemn this practice, but Dr Evgenii Sverdlov, the director of the institute, has resolved instead on a more pragmatic policy. "I want to understand what is in their minds", he says. So the departed are kept on the institute's books and even paid a portion of their salaries. "We cannot afford to make enemies", says Sverdlov.

Some also argue that the way in which junior colleagues, even students, have been compelled to take more responsibility on their shoulders is a blessing in disguise. All too often in the past, heads of laboratories regarded their subordinates as just that. Now there are bright-eyed students who play a part in the planning of research.

Even so, nobody can be sure that the supply of these young men and women will persist. A graduate student's life, never easy, is now one of hardship. The standard stipend, the equivalent of a few dollars a month, is plainly insufficient. Some students get by only because they live with their parents or are otherwise subsidized by them. At the elite schools, such as the Moscow Physical Technical Institute, the

undergraduate course lasts for six years.

The chances that these depressing prospects will feed back onto the age-groups feeding the better schools are, nevertheless, not high. One of Russia's most enduring assets may be the passion of those with bourgeois inclinations to see their sons and daughters educated well. Respect for intellectual achievement has survived the past 70 years well — too well, perhaps, for the health of tottering industry.

Meanwhile, it seems inevitable that basic research will need external help for some years to come. As things are, the most successful research groups are those that have established collaborations in the West. The benefits are huge: reagents otherwise unobtainable, pieces of equipment that could not be found locally, spare parts for machines that could not otherwise be used and occasional overseas trips. There is also, of course, a certain standing among competing groups that derives from such a collaboration. What stands out from a handful of working collaborations is that potential partners in the West are not driven by simple altruism. Partnerships emerge only when the Russian side can make a unique contribution. This period is flushing out islands of skill that make it seem worthwhile elsewhere to put up with the hassle of the still dreadful communications.

This is also the basis of the significant (if small) deals between companies in the West and Russian laboratories as different as the Kurchatov Institute in Moscow and the Electron Optics Institute at Omsk (in the southern Urals). Russian groups have contracted to carry out research in fields as different as plasma physics and fibre optics for the US Department of Energy and AT&T respectively. Monsanto and Sun Microsystems were even earlier in the field. Apart from the equipment and the modest travel these arrangements bring, these contracts pay salaries in US dollars. The going rate seems to have settled down at about \$60 a month — 7,500 rubles at the commercial rate of exchange.

Just what these numbers mean is a kind of mystery. At the end of last year, laboratories were talking of salaries of 1,800 rubles a month for senior people, essentially a doubling in four years. But since then, prices of some (not all) commodities have been 'liberalized'. In some areas, such as food-stuffs, some prices have increased by huge factors — bread, for example, by a factor of 50 or so — and, now, some have even begun to fall, as market economists say they should. No doubt they would fall faster, or rise less

quickly, if only Russia's chronically inefficient distribution system was reformed.

The worry is that there is still worse to come. Although the cost of a tankful of gasoline (petrol) is now ten per cent of a decent salary, the price is still less than that in the United States. (The Russian government is resisting demands by the International Monetary Fund that gasoline prices should rise to world levels as a condition for financial support.) Whole other areas of the economy, housing for example, remain an odd mixture of liberalization and the opposite. (People are encouraged to buy their own apartments, but cannot afford them, let alone the cost of maintenance.)

The publications part of the scientific enterprise has suffered most immediately from the economic upheavals chiefly because the price of paper has become realistic almost overnight. The editor of one science magazine, asked why he published so few book reviews, said, "There are no books". The publishing houses once owned by the state have been required to operate without losing money; many are doing that by hiving off the profitable parts of their businesses to collectives of their own employees.

Journals are still subsidized by the Russian Academy of Sciences, but, increasingly often, individual institutes at which particular journals are edited are making translation deals with publishers in the West, hoping that modest royalties paid in dollars will more than cover their ruble costs.

General circulation magazines have been much harder hit. The excellent general magazine whose Russian title translates as *Chemistry and Life* is a victim first of *glasnost* (which gave people such interesting other things to read that circulation began falling sharply from 300,000) and then of paper prices. (Circulation is now down to 120,000.) The magazine gets by with the help of a subsidy of 500,000 rubles a month from the Ministry of Information and Press.

Priroda (which means "Nature"), another academy magazine aimed at an audience such as that of *Scientific American*, lives more from hand to mouth. This journal, *Nature*, is doing what it can to help by trying to sell advertisements in the West, and plans to produce, the October issue.

Whether there will in due course be a market for these magazines, as for many other Russian products, is strictly conjectural. Perhaps because it is now summer, the optimists are in the ascendant.

John Maddox

Twenty-first century crops

Jim Peacock

GENETIC engineering has great potential in agriculture. By the turn of the century we will be using crop products which have been honed to market specification by the addition, subtraction or modification of genes. Transgenes will also be important in increasing the efficiency of crop production systems, and an exciting example of what can be achieved is given by Mariani *et al.* in this issue (*Nature* **357**, 384–387; 1992). They report on the restoration of pollen fertility in plants which have a genetically engineered pollen sterility system. The experimental plant is oilseed rape (*Brassica napus*), an important crop species, but the system is likely to be a general one which will find use in many crop and horticultural species — including maize.

The hybrid maize seed industry is large and is successful because farmers are prepared to buy F_1 hybrid seed each year. The yield of the hybrid, relative to the inbred parental lines, justifies the extra cost of purchasing hybrid seed. Hybrid corn also outperforms open-pollinated corn, a cheaper but more variable alternative. Hybrids are used in a number of other 'broadacre' crops (sunflower, sorghum) and in horticultural species (cabbages, tomatoes) because in each case the hybrid is again superior to open-pollinated lines. Although we still do not understand the molecular basis of hybrid vigour (heterosis), its exploitation is arguably the single most important contribution of genetic research to agriculture.

In maize, hybrid seed is produced by controlled pollination of a pollen-sterile female line. The male sterility of the female line is generated by manual or machine removal of the tassel, the structure bearing the anthers in which pollen development takes place. Until the late 1960s, hybrid corn was produced without this expensive and labour-intensive process. Instead, sterility was generated by a genetically determined system; DNA segment rearrangement in the mitochondrial genome resulted in abnormal gene expression during pollen development, preventing the formation of fertile pollen grains. Unfortunately, rearrangement of the mitochondrial genome was also associated with another phenotype, susceptibility to southern leaf blight, caused by the fungus *Helminthosporium maydis*. The widespread destruction of crops by this disease resulted in the disuse of the cytoplasmic male sterility system, and the consequent adoption of mechanical removal of the tassel.

For years geneticists have searched

without success for an alternative cytoplasmic male sterility system, or have attempted to develop nuclear gene systems. Now, genetic engineering has delivered a promising approach. Two years ago, Mariani *et al.* (*Nature* **347**, 737–741; 1990) described the consequences of introducing, into oilseed rape, a gene construct which coupled an anther-specific promoter to a bacterial coding sequence for a ribonuclease. This resulted in a spectacular, and specific, prevention of normal pollen development — the transgenic plant became male sterile. The promoter was from the *TA29* gene of tobacco, a gene expressed only in the tapetal cell layer of the developing anther. The ribonuclease, encoded by the *barnase* gene of *Bacillus amyloliquefaciens*, was cytotoxic, killing the tapetal cells and thus preventing pollen development.

This example of creative genetic engineering was particularly attractive because there was the possibility that a further component of the *Bacillus* ribonuclease system could be put to good

effect. In *Bacillus*, the ribonuclease is active extracellularly and the bacterium itself is protected by a protein coded by the *barstar* locus. The *barstar* product complexes with the ribonuclease and disables it.

In transgenic plants the *barstar* product, produced under instruction of the same *TA29* promoter, complexes with the coexpressed barnase molecule, neutralizing its cytotoxic properties and normal pollen development ensues. Mariani *et al.* now document the protein–protein interaction at genetic and cellular levels. A plant with the *barnase* gene is pollen sterile; a plant with the *barnase* and *barstar* genes is pollen fertile. Both genotypes are needed for the crossing system, thereby providing for the regular production of hybrid seed.

Here we have the powerful and novel heterologous application of a gene system, one which is exquisitely developed in the bacterial donor. Like it or not (and some do not) the genetic engineer is beginning to rival nature in adapting biological systems for specific purposes in our production of organisms. □

Jim Peacock is in the CSIRO Division of Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia.

BIOMINERALIZATION

Bacteria and the Midas touch

Stephen Mann

GOLD is the most precious of metals. How it comes to be concentrated in the stream beds of various regions around the world is a long-standing issue. The conventional view is that the placer deposits originate by mechanical weathering and transport of distant vein material. But John Watterson¹ now provides compelling evidence that many Alaskan gold nuggets arise from microbially induced chemical deposition. Although the involvement of bacteria in gold geochemistry is unlikely to set alight the City or initiate a biotechnological gold rush, it adds a little more glitter to the idea that microbiological processes are of fundamental importance in the cycling of metallic elements.

For some years now, bacterial activity has been implicated in the weathering, leaching and deposition of mineral ores. Microbes such as *Thiobacillus ferrooxidans* and *T. thiooxidans* are effective in leaching metals such as copper and uranium from insoluble minerals by redox processes at the mineral surface². Alternatively, bacteria can accumulate metals from solution by biosorption at the cell surface. In many cases, the metals are deposited as insoluble oxides

and sulphides, and large-scale and long-term activity can lead to important geological formations such as the Baltic Sea ferro-manganese concretions³. The possible influence of bacteria in gold geochemistry had been suggested^{4,5} but without much supporting field evidence. This is now rectified by Watterson¹, who presents striking evidence of lacelike, gold-decorated microbial structures associated with placer gold particles from nine Alaskan sites (Fig. 1).

What is even more exciting, is that the gold facsimiles are of such high resolution that there is a good preliminary evidence for the presence of specific bacteria — the *Pedomicrobium*-like budding bacteria. These bacteria are known to be involved in iron and manganese oxide deposition processes³, so it is a surprise to discover that they are also associated with gold accumulation. Scanning electron microscopy images suggest that the process of gold acquisition occurs during microbial growth and that this is sustained by the budding growth habit of *Pedomicrobium* which enables the colony to remain one step ahead of irrevocable entombment within its own gilded cage (Fig. 2). Although a mecha-

nical origin of placer gold is not ruled out, it is clear from Watterson's detailed work that most Alaskan placer gold particles are associated with microbial activity.

The microbial mediation of gold deposition from aqueous solution will come as no surprise to those involved in metal recovery or removal processes, as high levels of gold acquisition can be readily achieved using waste biomass. For example, algal mats of *Chlorella vulgaris* remove up to 90 per cent of the gold from dilute Au(III) solutions⁶. In previous work⁷, Watterson *et al.* showed that the spore coats of *Bacillus cereus* are active in nucleating metallic gold from dilute solutions of gold chloride. Indeed, it was these observations that led to the search for similar processes in the auriferous soils of Alaska.

It is generally considered that bacterial mineralization, unlike eukaryotic biomineralization, is the result of the adventitious interaction of extruded metabolic products (ions, gases, polypeptides, electrons) with extraneous metal ions in the surrounding environment (see table). Thus while complex, organized mineralized structures, such as coccoliths and diatom frustules, can be fabricated within intracellular vesicles of unicellular eukaryotes, the absence of spatially delineated compartments in bacteria limits the level of control exerted on nucleation and crystal growth. Correspondingly, the prodigious fine-tuning of crystal chemistry of eukaryotes is not generally encountered in bacterial mineralization; instead, the mineral products are located on or in the cell wall and are structurally ill-defined, physically heterogeneous and spatially disorganized. (There are a few contrary examples, notably the biosynthesis of magnetic minerals such as Fe₃O₄ and Fe₃S₄ in magnetotactic bacteria.)

Viewed from this general perspective,

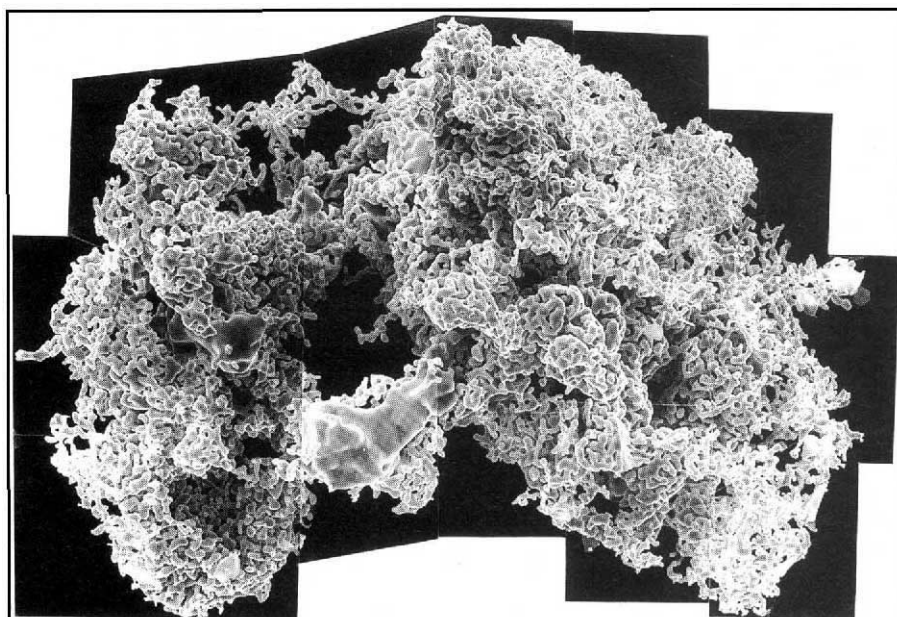


FIG.1 Scanning electron micrograph of a gold particle 140 μm by 100 μm , one of 18,000 grains recovered by J. Watterson from Lillian Creek, Alaska.

gold biometallization is an unusual, but intelligible example of bacterial biomineralization. As gold exists in solution usually in the form of Au(I) or Au(III) complexes — such as $[\text{Au}(\text{CN})_2]^-$, $[\text{Au}(\text{OH})(\text{Cl}_3)]^-$ and $[\text{AuCl}_4]^-$, many of which are easily reduced — non-specific electron transfer at the cell surface would be sufficient to induce metallic Au(0) deposition. Yet there are several aspects of the bacterial system which suggest some degree of specificity. First, the purity of the sequestered gold is very high — sometimes 24 carat. This implies that the redox and binding properties of the cell wall are selective for gold. In this regard, the predominance of *Pedomicrobium*-like budding bacteria in the Alaskan placer gold deposits is startling as they are commonly associated not with reductive

processes but with the oxidation of metals such as iron and manganese³. Interestingly, a dissimilatory bacterium, strain GS-15, induces the extracellular precipitation of magnetite⁸ (Fe₃O₄) or uranium (IV) oxide⁹ by reduction of Fe(III) or U(VI) present in the external medium. By comparison, it seems reasonable to suppose that gold complexes could also be used as terminal electron acceptors. If so, then this would be a novel form of energy transduction in anaerobic respiration.

Besides redox behaviour, other chemical differences between gold and iron or manganese suggest the need for some specificity in the process of gold biomineralization by *Pedomicrobium*-like bacteria. Sequestration of Fe(II) or Mn(II), as the cationic aquo complexes $[\text{Mn}(\text{H}_2\text{O}_6)]^{2+}$ and $[\text{Fe}(\text{OH})(\text{H}_2\text{O}_5)]^+$, is considered to be initiated by electrostatic binding to negatively charged acidic groups of bacterial cell-wall components such as the glutamate residues of peptidoglycans, phosphodiester linkages in teichoic acids and phosphate headgroups of lipopolysaccharides.

By contrast, stable water-soluble gold complexes are anionic, suggesting that positively charged residues (say RNH_3^+ , R_4N^+) or coordinating ligands such as RSH , R_2S , would facilitate gold sequestration. Indeed, *in vitro* binding of Au(I) to the alga, *Chlorella*, was shown by X-ray structure analysis (EXAFS) to involve two sulphur ligands¹⁰. Similarly, sulphur-rich keratin-like proteins are found in spore coats of *B. cereus*¹¹. So it seems feasible that specific polypeptides containing significant numbers of amino acids with sulphur ligands, such as cysteine, will be required to achieve the

BACTERIAL BIOMINERALIZATION PROCESSES

Process	Mechanism	Mineral	Examples
Biologically induced (biotransfer) (Extracellular)	Soluble biopolymers	Mn/FeOOH	<i>Leptothrix</i>
	Spore coats	MnOOH	<i>Pedomicrobium</i>
	Gas/ion exchange		<i>Bacillus</i>
	H ₂ S	Fe/CuS	<i>Desulfovibrio</i>
	CO ₂ /pH	CaCO ₃	<i>Calothrix</i>
	pH	MgNH ₄ PO ₄	<i>Proteus mirabilis</i>
	Proteolipid/ion transport	Ca ₅ (OH)(PO ₄) ₃	<i>Streptococcus</i>
	Phosphatase activity	(UO ₂) ₃ (PO ₄) ₂	<i>Citrobacter</i>
	Electron transfer	Fe ₃ O ₄	GS-15
		UO ₂	GS-15
Biologically controlled (biosynthesis) (Intracellular)	Nucleation proteins	H ₂ O(ice)	<i>Pseudomonas syringae</i>
	S-layer templates?	FeOOH	<i>Leptothrix</i>
		Fe ₃ O ₄	<i>Aquaspirillum</i>
			<i>magnetotacticum</i>
	Ferritin micelles	Fe ₃ S ₄ /FeS ₂	Wild types
		FeO(OH)/(PO ₄)	<i>P. aeruginosa</i>

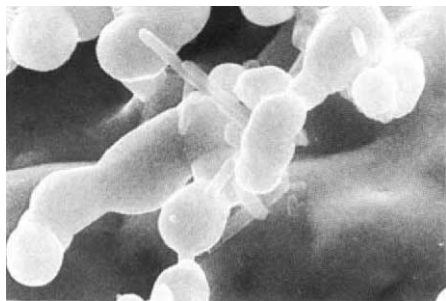


FIG. 2. Close-up of a gold particle, revealing the rod-like hyphae, 0.2 μm in diameter, through which new bacterial cells bud.

high localized levels of bacterial gold accumulation seen in the Alaskan placer deposits. Moreover, it is possible that aggregates of these polypeptides might act as nucleation sites in the cell walls, as observed, for example, in bacteria capable of nucleating ice¹².

Clearly, there is much to do. A systematic study of gold deposition by *Pedomicrobium* cultures, in the presence and absence of iron and manganese, would be very interesting. Isolation and identification of the active constituents could lead to some intriguing experiments. One is reminded of the successful cloning and expression, for frost-proofing plants, of the ice-nucleating proteins of *Escherichia coli*¹³. Is there an analogous 'Midas-gene', from which gold-nucleating *E. coli* could be engineered? More down to earth, Watterson cites preliminary evidence for the presence of budding bacteria in Precambrian Witwatersrand gold deposits, from South Africa, based on previously published but contentious electron micrographs¹⁴. More extensive and detailed studies are required, but it seems that microbial involvement in gold geochemistry may reach back over 2,800 million years. The quest for bacterioform gold has begun. □

Stephen Mann is in the School of Chemistry, University of Bath, Bath BA2 7AY, UK.

The work that muscles can do

R. McNeill Alexander

BY choosing a particularly suitable subject, Marsh and his colleagues (page 411 of this issue¹) have been able to measure a muscle's natural performance. That is something of an achievement — experiments with isolated muscles have told us how strongly and how quickly many different muscles can contract, but it is much more difficult to discover how muscles perform in natural movements of living animals.

The subject used by Marsh *et al.* was the scallop, a bivalve mollusc that swims by clapping the two valves of its shell open and shut: a single joint is worked by a single muscle so there is no ambiguity about the muscle's contribution. Marsh and his colleagues attached transducers to free-swimming scallops, to record the movements of their valves and the pressure inside them. From the records they calculated the length changes of the muscle, the force it exerted and the work it did in each contraction. The significance of this type of work was highlighted by A. V. Hill in a classic paper². The forces that different muscles can exert are proportional to their cross-sectional areas, and the distances they can shorten are proportional to their lengths. Thus the work (force multiplied by shortening distance) done in a contraction is proportional to muscle volume. Different muscles have different intrinsic speeds and are capable of very different power outputs, but the work done per unit volume (or per unit mass) of muscle, in a single contraction, should be about the same for all muscles of similar composition.

How much should this work be? Vertebrate striated muscles, shortening very slowly, exert stresses up to about $3 \times 10^5 \text{ N m}^{-2}$ (ref. 3). Other muscles with myofilaments of similar length can be expected to do about the same (but some invertebrate muscles with longer myofilaments can exert larger stresses). Muscles seem typically to shorten by 25% or less in normal movements (if they shortened much more, the stresses they could exert near the extremes of the length range would be greatly reduced). The density of muscle is about $1,060 \text{ kg m}^{-3}$. Thus muscles can be expected to do about $3 \times 10^5 \times 0.25/1,060 = 70 \text{ J kg}^{-1}$ in a strong, slow contraction.

To maximize power output, the muscle must contract faster, at a rate at which the stress is reduced by a factor of 0.3 (ref. 3), making the work output only 20 J kg^{-1} . Thus we may expect vertebrate and other similar muscles to do 20 J kg^{-1} in each contraction when maximum power output is required, and

up to 70 J kg^{-1} when the quantity of work from a single contraction is more important than the rate at which the work is done. Work-loop experiments with isolated fibre bundles have usually given values lower than 20 J kg^{-1} , although R. S. James (personal communication) finds values of up to 23 J kg^{-1} for the mouse extensor digitorum longus.

The scallop muscle investigated by Marsh *et al.* has myofilament lengths similar to those of vertebrate muscle and had a work output of up to 20 J kg^{-1} in each contraction. That is about what we should expect of a muscle contracting at the rate that maximizes power output. The best other data for muscle performance in intact animals come from wallabies and starlings with surgically implanted transducers. Griffiths⁴ found that when a wallaby accelerated from rest the medial gastrocnemius muscle did about 30 J kg^{-1} (from his Fig. 6D, using my own measurement of the mass of the muscle in a closely related species). Biewener and colleagues⁵ attached strain gauges to the humerus of starlings which they then flew in a wind tunnel. They used films to record wing movements and were able to calculate that the work done by the muscle in each wing beat was only 7 J kg^{-1} .

These values contrast with much higher estimated work outputs in some more strenuous activities in which (unfortunately) the contributions of individual muscles have not been measured. The highest recorded jump of a bushbaby (*Galago*)⁶, to an amazing 2.3 m, must have been powered by the leg and back muscles which are 24% of body mass⁷ and must have done 88 J kg^{-1} . Sprint cyclists reach up to 560 J for each revolution of the pedals⁸ using leg muscles totalling (probably) about 14 kg : this is 40 J kg^{-1} . Hawk wing muscles have been estimated at 50 J kg^{-1} for each wing beat of a fast climb⁹. The squid *Illex* jet propels with pressures up to 50 kN m^{-2} , driving out water amounting to 45% of

1. Watterson, J. R. *Geology* **20**, 315–318 (1992).
2. Briertley, J. A. & Briertley, C. L. in *Biomineralization and Biological Metal Accumulation* (eds Westbroek, P. & de Jong, E. W.) 499–509 (Reidel, Dordrecht, 1983).
3. Ghiorse, W. C. & Hirsch, P. *Appl. Envir. Microbiol.* **43**, 1464–1472 (1982).
4. Lakin, H. W. *et al. Bull. U.S. Geol. Surv.* **1330**, 80 (1974).
5. Beveridge, T. J. *et al. Appl. Envir. Microbiol.* **45**, 1094–1108 (1983).
6. Darnall, D. W. *et al. Envir. Sci. Technol.* **20**, 206–208 (1986).
7. Watterson, J. R., Nishi, J. M. & Botinelly, T. *U.S. Geol. Surv. Circ.* **945**, 1–5 (1985).
8. Lovley, D. R. *et al. Nature* **330**, 252–254 (1987).
9. Lovley, D. R. *et al. Nature* **350**, 413–416 (1991).
10. Watkins, J. W. *et al. Inorg. Chem.* **26**, 1635 (1987).
11. Aaronson, A. L. & Pandey, N. K. in *Spores VII* (eds Chambliss, G. & Vary, J. C.) 54–61 (Am. Soc. Microbiol., Iowa, 1977).
12. Warren, G. J. & Wolber, P. K. *Cryo-Lett.* **8**, 204–215 (1987).
13. Wolber, P. K. *et al. Proc. natl. Acad. Sci. U.S.A.* **83**, 7256–7260 (1986).
14. Hallbauer, D. K. & van Warmelo, K. T. *Precamb. Res.* **1**, 199–212 (1974).

1. Marsh, R. L., Olson, J. M. & Guzik, S. K. *Nature* **357**, 411–413 (1992).
2. Hill, A. V. *Sci. Progr.* **38**, 209–230 (1950).
3. Woledge, R. C., Curtin, N. A. & Homsher, E. *Energetic Aspects of Muscle Contraction* (Academic, London, 1985).
4. Griffiths, R. I. *J. exp. Biol.* **147**, 439–456 (1989).
5. Biewener, A. A., Dial, K. P. & Goslow, G. E. *J. exp. Biol.* **164**, 1–18 (1992).
6. Hall-Craggs, E. C. B. *J. Zool., Lond.* **147**, 20–29 (1965).
7. Grand, T. I. *Am. J. phys. Anthropol.* **47**, 211–240 (1977).
8. Davies, C. T. M. & Sandstrom, E. R. *Eur. J. appl. Physiol.* **58**, 838–844 (1989).
9. Pennycuik, C. J., Fuller, M. R. & McAllister, L. J. *J. exp. Biol.* **142**, 17–29 (1989).
10. O'Dor, R. K. & Webber, D. M. *J. exp. Biol.* **160**, 93–112 (1991).

body volume¹⁰. To achieve this, its mantle muscles (40% of body mass) must put out more than 50 J kg^{-1} of work. This performance can reasonably be compared with that of the vertebrate and scallop muscles, as all these muscles have similar myofilament lengths, but not with the well-known data for insect jumping (the insect muscles have longer myofilaments).

We might expect to find muscles doing 20 J kg^{-1} in activities that require maximum power (for example, swimming or cycling) and 70 J kg^{-1} when maximum

work is required from a single contraction (as in jumping). By these standards, some of the work outputs recorded so far seem remarkably high, but none of the high values came from measurements of the output of individual muscles. We need more measurements like those of Marsh and his colleagues to give us a general picture of how muscles perform in strenuous activities. □

R. McNeill Alexander is in the Department of Pure and Applied Biology, University of Leeds, Leeds LS2 9JT, UK.

ASTRONOMY

Fire burn and cauldron bubble

Leo Blitz

THE disk of the Milky Way is thought to be peppered with explosions that clear out vast volumes of space which then expand and burst in the halo of the Galaxy. Three Berkeley astronomers believe that they have found evidence of the ruptured shells in a new analysis of archival observations of the dust and gas in the disk of the Milky Way (B.-C. Koo, C. Heiles & W. T. Reach *Astrophys. J.* **390**, 108–132; 1992). The analysis finds “worm-like” structures which are most plausibly explained as the remnants of the bubbles of hot gas that have burst and spewn their hot interiors into the halo of the Milky Way.

The remnants of exploding stars, or supernovae, are easily identified by radio astronomers up to a million years after the explosion. Their distinct signature is non-thermal radiation generated by relativistic electrons spiralling about the magnetic fields compressed by the explosion itself. Although both young and old stars can give rise to supernovae, the most massive stars, none of which is thought to live more than several million years, are thought always to end their lives this way. These ‘O’ stars are frequently formed in clusters in a time span short compared to their lifetimes.

During this time, the stars do not move far from their birthplaces and will explode close to where they were born. The photograph shows such a young cluster, NGC 2244, the ultraviolet radiation from which excites the Rosette Nebula, the beautiful, symmetric emission nebula surrounding the stars. The five brightest stars in the photo are O stars and the central hole in the nebula is a cavity apparently evacuated by energetic winds emanating from them. The cavity is expanding at a velocity of 20 km s^{-1} and at the present rate will double its size in about 200,000 years.

By the time the cavity is three to five times its current size, the first of these

stars will have ended its life in a supernova explosion, inflating the cavity, and heating the gas in its interior. Each of the bright, massive stars will subsequently explode, further heating the bubble to a million kelvin or higher. The outer envelope will plough into the ambient gas, sweeping up the neutral interstellar gas into a thick expanding shell. At this stage, the structure is a supershell, a term coined by Carl Heiles for structures he found throughout the disk of atomic hydrogen in the Milky Way.

What happens next is rather murky. Neighbouring shells might merge to form a network of tunnels of hot gas in the disk. Some theorists have even proposed that most of the volume of the disk is

actually hot gas from overlapping supernova remnants. Another idea is that the superbubbles continue to expand in the gaseous disk of the Milky Way until their diameters are comparable to the thickness of the atomic hydrogen layer. At this stage, the supershells experience a greatly reduced ambient pressure at the points farthest from the galactic mid-plane, and expand rapidly into the galactic halo until the bubble itself bursts.

Astronomers have proposed that this may be the source of the tenuous hot gas in the halo which is detected in the ultraviolet and X-ray regions of the spectrum. The breached bubbles may also be the source of galactic fountains of hot gas shot up into the halo which subsequently cool and rain back down onto the galactic plane. Theorists have termed the structures that are expected from the burst bubbles ‘chimneys’, not only because they would channel hot gas from the disk, but because the heavy-element-enriched gas would pollute the relatively pristine environment of the halo.

Until now, good observational evidence for the burst or bursting bubbles in the Milky Way has been rather scant. The problem is that one must view these features through the disk of the Galaxy and they would appear only as slight increases over the totality of emission from the disk. What Koo, Heiles and Reach have done is to apply a technique known as median-filtering to enhance the low-surface-brightness features. A map is produced by taking the median value of the emission over some designated area, which is the filter size, and



The Rosette Nebula and the star cluster NGC2244 (courtesy of David L. Block, Claus Madsen and the Schmidt Telescope Unit, La Silla, Chile).

subtracting it from the data. Objects smaller than the filter size, the equivalent of an electronic high-pass filter, are highlighted in the resulting image.

The technique was applied to three surveys of the plane of the Milky Way: a 21-cm radio survey of atomic hydrogen, a 100- μ m microwave survey of the cool dust, and a 408 MHz survey of radio continuum emission, which comes primarily from non-thermal synchrotron processes. The maps show 118 striking vertical structures that Heiles has previously termed worms, because they are wiggly things that look as if they are crawling out of the galactic plane. Worms are the observational analogue of the chimneys; each worm has not necessarily been shown to vent into the galactic halo, the *sine qua non* for a chimney. The most noteworthy feature of the analysis is that 35 of the worms are correlated in all three maps. Because atomic hydrogen emits a discrete spectral line, the Doppler velocity combined with the known rotation law of the Milky Way makes it possible to determine a distance, and consequently the physical properties of the worms.

The authors analyse the energetic requirements of the worms and show that multiple supernovae, presumably from clumps of O stars similar to those in the Rosette Nebula are required to explain the observed emission. Almost half of the worms are associated with H II (ionized hydrogen) regions, the class of objects like the Rosette, which is what one would expect if the worms are produced by clustered supernovae. Their measured sizes are just what one would expect from superbubbles expanding and bursting into the halo. Furthermore, a correlation between emission by warm dust and the 408 MHz emission suggests that much of the radio emission from the worms is thermal, a surprising conclusion, one which further suggests that the walls of the burst bubbles are partially ionized by the O stars that remain unseen in the cluster.

The work has yet another important ramification for gas in the interstellar medium. It has been known for some years that there is diffuse, warm (10,000 K), ionized gas in the interstellar medium. Ron Reynolds (University of Wisconsin) has been largely responsible for showing how pervasive this component is in the disk of the Milky Way. However, the very pervasiveness has caused a puzzle: what keeps it all so warm? In a relatively uniform interstellar medium, the ultraviolet photons could not penetrate to the far reaches of the interstellar medium because they would be absorbed near the O stars from which they are generated.

Heiles, however, has argued that the warm, ionized gas throughout the

Galaxy comes from the walls of worms and chimneys: the ionizing photons traverse large path lengths through rarified gas before being absorbed by the walls of the bubbles cleared out by the supernova explosions. Thus an outstanding problem in the energy distribution of the interstellar gas has been solved if it can be confirmed by other observations.

In the 25 years since Nigel Calder wrote the *Violent Universe*, our conception of the Universe has changed from a gentle place where events tick along with clockwork simplicity to a place domin-

ated by explosive events of gargantuan ferocity. Similarly, our view of the gaseous disk of the Milky Way has changed from one where energetic events like supernovae are incidental to its workings. Now we see that their violence, amplified by the clustering of the detonating stars, dominates the structure and evolution of the interstellar medium. $\square$

Leo Blitz is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA.

HUMAN EVOLUTION

Diabetes running wild

Jared M. Diamond

STUDIES of a diabetes epidemic on a Pacific islet, carried out by Paul Zimmet and colleagues¹⁻⁶, will make grim reading for public health professionals, geneticists and students of human evolution. Among older adults, the prevalence of diabetes on Nauru Island is over 60%, a horrifying figure which may be but a foretaste of a sharp rise in the incidence of diabetes in other developing countries. Events on Nauru may also illustrate natural selection in action, shifting human gene frequencies, while exemplifying the paradox of an epidemic of a non-infectious genetic disease.

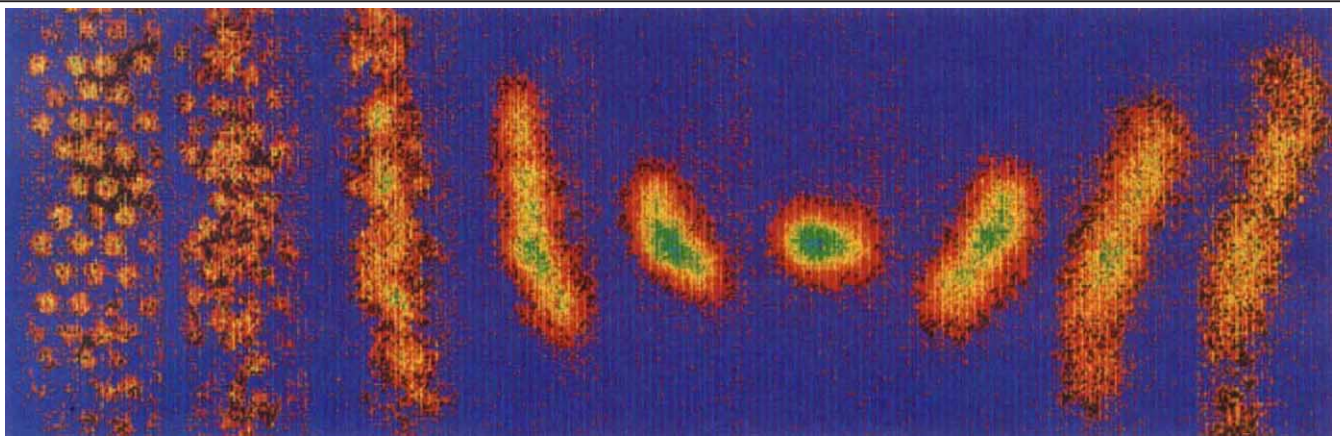
At the outset, let us be clear about the differences between type II diabetes (NIDDM, non-insulin-dependent diabetes mellitus, the subject of this article) and the much less common type I diabetes (IDDM). The first differs in its genetic basis, age of onset (mainly in adults instead of juveniles), cellular manifestations (peripheral insulin resistance rather than autoimmune destruction of pancreatic islet β -cells), and treatment (often by diet, not by insulin injections)⁷. It has strong genetic and environmental components, the environmental risk factors including the hallmarks of a modern Western lifestyle: low physical activity, high calorie intake of a supermarket diet, and obesity¹. It is the switch to such a lifestyle ('cocolonization')³ accompanying rising affluence that is now precipitating NIDDM epidemics in many parts of the developing world¹⁻⁶.

One of the most serious of these epidemics is on Nauru, a remote atoll occupied by 5,000 Micronesians whose formerly energetic lifestyle depended on fishing and subsistence farming. Colonization by Britain, Australia and New Zealand, and income from phosphate mining, transformed Nauruans into one of the world's wealthiest, most sedentary peoples. Virtually all food is now im-

ported and energy-dense; calorie intake is more than double Australian recommended norms; and obesity is rampant. NIDDM used to be nonexistent but a severe form striking many young adults reached epidemic frequencies after 1950 and now affects almost two-thirds of adults by age 55-64. The disease now contributes to most non-accidental deaths on Nauru, with the paradoxical result that wealthy Nauru has one of the world's shortest human lifespans⁵.

Analyses² of health surveys conducted on Nauru in 1975-76, 1982 and 1987 trace the course of the epidemic. Although the age-standardized prevalence of NIDDM had increased from practically 0% to 28% in the preceding decades, the corresponding prevalence of impaired glucose tolerance (first stage in development of NIDDM) actually declined from 21% in 1975-76 to 9% in 1987, while the incidence of progression from normal health to impaired glucose tolerance or full-blown NIDDM more than halved in those same years. As a result, the age-standardized prevalence of NIDDM in 1987 had dropped below 1975-76 levels.

Evidently, the NIDDM epidemic on Nauru has passed its peak. But this outcome cannot be attributed to the most obvious explanation, a decline in environmental risk factors, because they remain at least as severe as before. Rather, it seems that NIDDM associated with the new lifestyle has already struck most of the genetically susceptible Nauruans, leaving the unaffected population to consist mostly of genetically resistant individuals. Nauru's epidemic of a genetic/environmental disease thus resembles an infectious disease epidemic running its course, but the decline in disease frequency stems from an entirely different mechanism: inter-individual variation in genetic susceptibility, rather than acquired immunity. It illustrates



NEUTRON beams have found a variety of applications, from cancer therapy to investigations of atomic structures. But neutrons are recalcitrant creatures: they cannot be steered by electromagnetic fields, and the weakness of their interaction with ordinary matter — which is what makes them such penetrating probes — means that they are hard to reflect or refract. Nevertheless, the fuzzy patches shown here are images of a neutron beam focused by, of all things, a glass

lens — one made of tiny hollow capillaries bundled together. M. A. Kumakhov and colleagues show on pages 390 and 391 that neutrons are channelled inside curved capillaries by internal reflection, and that a stack of capillaries can be used to steer neutrons at moderate curvature or, as here, to bring them to a common focus. The central frame shows neutrons at the focal point of the lens, and the extended patterns show the beam profile either side of the focus. D.L.

natural selection operating on humans under our very eyes.

This conclusion may surprise those who are accustomed to thinking of NIDDM as a disease without selective consequences because it supposedly strikes 'post-reproductive' adults. In traditional human societies, however, no adult ever becomes post-reproductive in a broad sense: adults over 50 remain crucial to the survival of their children and grandchildren. Even in a narrower sense, NIDDM is not a post-reproductive disease on Nauru, where it often strikes at peak reproductive age. Diabetic women on Nauru have more stillbirths and less than half as many live births as healthy controls². Through natural selection, these narrow and broad effects would cause a diabetogenic genotype to decline in frequency in a diabetes-eliciting environment, as actually observed in successive generations of laboratory rats.

The diabetes epidemic on Nauru is not just a freak event, but a harbinger of events in progress throughout much of the developing world. Epidemics of NIDDM accompanying coca-colonization have already been noted among groups of Polynesians, American Indians and Aboriginal Australians^{3,5}. Ominously, they have also been reported among coca-colonized Chinese and Asian Indian emigrants to Mauritius, where the current NIDDM prevalence of over 20% in the Chinese contrasts with that of 1% in mainland China⁴. A sharp rise in the frequency of NIDDM also emerges from studies of Asian Indian emigrants to Fiji, South Africa and Britain, and of Chinese in Singapore, Taiwan and Hong Kong. As lifestyles change in China and India themselves, by far the most populous

countries in the world, NIDDM may create an enormous public health problem; Zimmet⁶ estimates there will be 50 million diabetics in these two nations by the year 2000.

Because the now-lethal NIDDM genotype somehow became widespread in many ethnic groups under their formerly spartan living conditions, what advantage did the genotype provide then? The most plausible interpretation is Neel's 'thrifty genotype' hypothesis⁸. According to this view, under the conditions of fluctuating and usually meagre food supply that prevailed throughout most of human history, those individuals with 'thrifty' adaptations such as hair-triggered insulin release would convert more of their ingested calories into fat during occasional food gluts. They would thus be better able to survive bouts of starvation, and would develop diabetes only under modern conditions of constantly available high-calorie food and little exercise. This view is supported by the frequent development of diabetes in 'affluent' zoo-housed monkeys, and by the better ability of diabetic rats to survive periods of starvation.

Although most human populations would have been selected in this way for a thrifty genotype, selection would have been especially intense on Nauruans for three reasons: their ancestors reached the island by long canoe voyages on which initially fatter people would have been more likely to escape death by starvation; droughts and crop failures were common on Nauru in times past; and most of the population was forcibly deported (and over one-quarter died of starvation) during the Japanese occupation of 1942–45². Because Nauruans used to be more starvation-prone than

most people, and have now become wealthier faster, the frequency with which they suffer from NIDDM is now especially high. But they differ from the rest of humanity only by degree.

One anomaly remains: why, when NIDDM is striking so many peoples, is its incidence so low in whites of the industrialized nations (for example only 8% in the United States)? As an explanation, the Nauru studies point to natural selection operating in the past². We term our modern, sedentary, high-calorie, overweight lifestyle a Western lifestyle precisely because it first arose in Western industrialized nations. Before modern medicine made NIDDM more manageable, genetically susceptible Europeans would have been gradually eliminated, bringing NIDDM to its present low frequency. If so, the Nauru epidemic gains additional interest, by telescoping into a few coca-colonized generations the operation of natural selection for so many centuries in Europe. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA. His book *The Rise and Fall of the Third Chimpanzee* has won the 1992 Rhône-Poulenc science book prize, as announced last week.

1. Zimmet, P. *Diabetes Care* **15**, 232–252 (1992).
2. Dowse, G. K. et al. *Am. J. Epidemiol.* **133**, 1093–1104 (1991).
3. Zimmet, P. *Diabetes Annual* **6**, 1–19 (1991).
4. Dowse, G. K. et al. *Diabetes* **39**, 390–396 (1990).
5. Zimmet, P. et al. *Diabetes/Metabolism Reviews* **6**, 91–124 (1990).
6. Zimmet, P. *Int. Diabetes Fdn Bull.* **36**, 30–32 (1991).
7. Vadheim, C. et al. in *Principles and Practice of Medical Genetics*, Vol. 2 (eds Emery, A. & Rimoin, D.) 1521–1558 (Churchill Livingstone, Edinburgh, 1990).
8. Neel, J. V. *Am. J. hum. Genet.* **14**, 353–362 (1962).

Dead Issues

THE idea that some mass extinctions were caused by massive volcanic episodes is undermined by D. H. Erwin and T. A. Vogel in *Geophysical Research Letters* (**19**, 893–896; 1992). The greatest extinction, the Permo/Triassic, has been linked to volcanics in southern China with an estimated volume of 1,000 km³. For comparison, Krakatau, which erupted just 12 km³ of magma and raised ash to a height of 30 km, caused global cooling of 0.4 °C for two years. The authors test the idea by reversing it, looking at some of the greatest volcanic episodes (the largest, at Elkhorn Mountains 75–81 million years ago, erupted at least 4,000 km³ of magma) to see if they coincide with increases over the natural extinction rate. But in none do the authors find any related decrease in biological diversity, even regionally.

Molecular lightning rod

THE pursuit of the magic bullet that will penetrate a cell and strike only its target protein was probably started by Ehrlich. One version depends on bringing a chromophore close to the target protein and exciting it with laser light, thus sensitizing the photochemical destruction of the protein. The dye can be coupled to an antibody or dispersed in a membrane or micelle if the target is a membrane protein. K. G. Linden *et al.* (*Biophys. J.* **61**, 956–962; 1992) have now defined the distance over which the excited chromophore will exert its effect: a labelled antibody, attached to its antigen, β -galactosidase, causes (under given conditions) 75% inactivation of the enzyme. Attached to protein A, which is then bound to the antibody, it destroys 20% of the activity and on anti-protein A, bound to protein A, itself bound to the enzyme-associated anti- β -galactosidase, it can annihilate no more than 4% of the activity.

Cereal success

FOR all its enormous economic importance, wheat has proved a tough subject for the genetic engineer. But the door to the manipulation of its genome springs ajar with a report by V. Vasil *et al.* (*BioTechnology* **10**, 667–674; 1992) of the production of fertile transgenic plants. The gene involved (*bar*) is selectable because it encodes an enzyme which inactivates the main agent of a broad-spectrum herbicide; it was delivered, in plasmid form, into the cells of an embryogenic callus by microprojectile bombardment. The resulting plants exhibited *bar* activity down to the third generation, and the gene segregated in dominant mendelian form. The success rate of gene transfer was low, but the door to the engineering of drought- and pest-resistant wheat will no doubt be pushed open further in the coming months.

Transcriptional transgressions

Michael R. Green

EUKARYOTIC RNA polymerases, the enzymes that transcribe the genetic information, are unable to initiate the process accurately by themselves; instead, they are directed to the start-site of a gene by a group of auxiliary proteins, dubbed transcription factors. Over the past 15 years some general principles about transcription factors have emerged — or rather we thought they had emerged, for a series of papers^{1–5} has now brought three of those principles into question.

Eukaryotic genes can be divided into three classes, according to which RNA polymerase is responsible for transcribing them: thus RNA polymerase I transcribes ribosomal RNA (class I) genes; RNA polymerase II transcribes mRNA-encoding (class II) genes; and RNA polymerase III transcribes class III genes, which encode small RNAs such as 5S RNA and transfer RNAs. The transcription factors for each RNA polymerase were originally defined by chromatographic fractionation.

First principle

One principle to emerge was that each RNA polymerase has its own set of transcription factors. Yet several observations revolving around the transcription factor TFIID have blurred the demarcation between transcription factor classes. TFIID, a multiprotein complex, is the prototype class II transcription factor. It is best known for binding to the TATA box, the major landmark of a minimal (or core) class II promoter. But some class III genes (for example U6 snRNA, 7SK and EBER2) contain an upstream TATA-like element required for full transcriptional activity. Biochemical complementation experiments have verified that the TATA-box binding protein (TBP), a subunit of TFIID, is indeed required for transcription of these class III genes^{6–8}.

It had been generally assumed, however, that the TATA-containing class III promoters were quirks; the more conventional class III genes (tRNA and 5S RNA, for example) have completely intragenic promoters and lack an upstream TATA box element. Yet White and colleagues² have used both biochemical complementation and TATA box oligonucleotide-inhibition experiments to show that TBP is also required for the transcription of these typical class III genes.

But the TBP story does not end here. Comai *et al.*³ demonstrate that TBP is also involved in RNA polymerase I-directed transcription; specifically, anti-

TBP antibodies can deplete the RNA polymerase I transcription factor SL1, thereby abolishing transcription. RNA polymerase I transcription was restored by addition of SL1 but not TBP. This is because SL1 is not just TBP, but rather a complex of TBP and three TBP-associated factors (TAFs). TFIID is also a complex of TBP and several TAFs, but the constellations of TAFs in TFIID and SL1 are completely different³.

Although the TBP-containing complex involved in RNA polymerase III transcription has not been characterized, several observations suggest that it is TFIIB, a well-known RNA polymerase III transcription factor. The unifying concept taking shape is that all three RNA polymerases require an essential transcription factor, which is a complex containing TBP and TAFs, the composition of the TAFs differing in the three instances.

These findings explain why TBP does not behave as a single biochemical entity. For example, it is present both in the conventional TFIID fraction and in another abundant complex, designated B-TFIID (ref. 9). It now seems likely that B-TFIID is involved in either class I or class III transcription.

Surprisingly, however, B-TFIID can substitute for the RNA polymerase II transcription factor, TFIID, in a basal transcription reaction⁹. B-TFIID can therefore provide the TBP function of TFIID. Minimally, this means that TBP in B-TFIID can contact the TATA box and several RNA polymerase II trans-

1. Parvin, J. D., Timmers, H. T. M. & Sharp, P. A. *Cell* **68**, 1135–1144 (1992).
2. White, R. J., Jackson, S. P. & Rigby, P. W. J. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1949–1953 (1992).
3. Comai, L., Tanese, N. & Tjian, R. *Cell* **68**, 965–976 (1992).
4. Roy, A. L., Meisterernst, M., Pognonec, P. & Roeder, R. G. *Nature* **354**, 245–248 (1991).
5. Sumimoto, H., Okhuma, Y., Tamamoto, T., Horikoshi, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9158–9162 (1990).
6. Lobo, S. M., Lister, J., Sullivan, M. L. & Hernandez, N. *Genes Dev.* **5**, 1477–1489 (1991).
7. Margottin, F. *et al.* *Science* **251**, 424–426 (1991).
8. Simmen, K. A. *et al.* *EMBO J.* **10**, 1853–1862 (1991).
9. Timmers, H. T. M. & Sharp, P. A. *Genes Dev.* **5**, 1946–1956 (1991).
10. Stringer, K. F., Ingles, C. J. & Greenblatt, J. *Nature* **345**, 783–786 (1990).
11. Horikoshi, N. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 5124–5128 (1991).
12. Lee, W. S., Cheng-Kao, C., Bryant, G. O., Liu, X. & Berk, A. J. *Cell* **67**, 365–376 (1991).
13. Liberman, P. M. & Berk, A. J. *Genes Dev.* **5**, 2441–2454 (1991).
14. Seto, E., Shi, Y. & Shen, T. *Nature* **354**, 241–245 (1991).
15. Okhuma, Y., Sumimoto, H., Horikoshi, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9163–9167 (1990).
16. Vaishnav, Y. N. & Wong-Staal, F. A. *Rev. Biochem.* **60**, 577–630 (1991).
17. Tamura, T., Sumita, K., Hirose, S. & Mikoshiba, K. *EMBO J.* **9**, 3101–3108 (1990).
18. McCormick, A., Brady, H., Fukushima, J. & Karin, M. *Genes Dev.* **5**, 1490–1503 (1991).

cription factors, such as TFIIA and TFIIB. Thus, these surfaces of TBP are accessible (or potentially accessible) in the B-TFIID complex despite the fact that TBP is enshrouded in an ensemble of TAFs irrelevant to RNA polymerase II function.

Taken together, these studies suggest that TBP can interact with DNA, three different sets of TAFs, and several RNA polymerase II transcription factors. To complicate matters further are reports that TBP can be directly contacted by several transcriptional activators including VP16 (ref. 10), E1a (refs 11, 12) and ZTa (ref. 13). The many contacts imposed on TBP by these observations are rather a tall order for a protein of a relative molecular mass of only 38,000. Perhaps the TBP-TAF arrangement is sufficiently dynamic that TAFs can be displaced from TBP to accommodate contacts by other general transcription factors and activators.

The new findings of course themselves raise a series of questions. What, for instance, does TBP do in the class I and class III transcription factor complexes? Because the promoters of class I and conventional class III genes lack TATA elements, it is not clear that the function of TBP in these complexes is DNA binding. Perhaps, as discussed above, TBP is particularly effective at providing a nidus for protein-protein interactions. A further question concerns the total number of TBP-containing complexes. Are there several complexes, which differ in their composition of TAFs, for a given RNA polymerase?

Second principle

A second principle, also now in some jeopardy, is that each transcription factor performs a unique function. This notion has been challenged most severely by studies on the RNA polymerase II transcription factor, TFIIA. TFIIA has always been mired in controversy: simply put, not all investigators could find a requirement for it. The simplest explanation for this discrepancy was that TFIIA was also present in one of the other fractions used for *in vitro* transcription. But work from Roeder and colleagues now points to a more subtle explanation. They have described a factor, designated TFIIG, that is chromatographically unrelated to TFIIA but can functionally substitute for it⁵. More recently, this laboratory has described another factor, TFII-I, which binds to the transcription start site (initiator element) and can also take the place of TFIIA⁴. Thus, three apparently distinct factors, TFIIA, TFIIG and TFII-I, can perform a similar function. The discovery of other initiator-binding proteins, such as YY1 (ref. 14), implies that additional redundancy is likely to be unearthed.

Third principle

A third principle to have emerged is that each transcription factor performs a 'universal' function, required for all promoters of that class. This idea has been challenged by work on the immunoglobulin heavy-chain gene (IgH) promoter. Parvin *et al.*¹ report that the general transcription factor TFIIE, required for transcription of the adenovirus (Ad) ML promoter, is not needed for transcription of the IgH promoter. This conclusion stems both from the lack of a requirement for TFIIE in a complementation assay, and the failure of an anti-TFIIE antibody to inhibit transcription from the IgH promoter.

TFIIE is not just required for the Ad ML promoter, but also for other promoters such as H2B and H4 (ref. 15). What could be the basis for a promoter-specific requirement of a basal transcription factor? The simplest possibility is that, for the IgH promoter, another factor supplies TFIIE function. This putative factor could, for instance, interact with the core IgH promoter in a sequence-specific fashion. A precedent for this idea is that core promoters can harbour binding sites for regulatory proteins. For example, the region encompassing the core promoter of the human immunodeficiency virus contains binding

sites for several factors (see ref. 16), and the core promoters of the myelin basic protein¹⁷ and the pituitary transcription factor, GHF1 (ref. 18) are able to direct tissue-specific expression.

A second, more radical, possibility is that transcription of the IgH promoter does not require TFIIE activity. Although TFIIE has been purified and cloned, its precise function remains unknown, making this notion difficult to evaluate. A final possibility, which cannot yet be ruled out, is that TFIIE is, after all, required for transcription of the IgH promoter, but that the level of TFIIE required is much lower than that needed for the Ad ML promoter.

Whichever explanation is correct, as Parvin *et al.*¹ point out their results raise a potential problem with nomenclature. RNA polymerase II transcription factors have been referred to as general or basic, the two terms being used interchangeably. The study of the IgH promoter suggests that some of these factors may be basic but not general. And some previous generalizations in transcription may be basically untrue. □

Michael R. Green is in the Program in Molecular Medicine, University of Massachusetts Medical Center, Worcester, Massachusetts 01605, USA.

MATERIALS SCIENCE

Strength in disunity

Paul Calvert

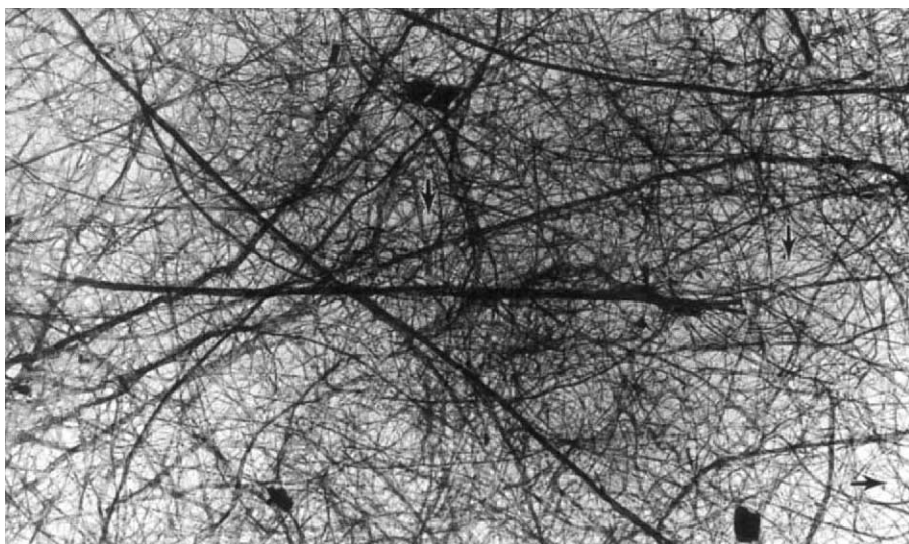
REINFORCED polymer composites are increasingly used in the aerospace industry because they offer an unrivalled combination of low density with stiffness and strength. The archetypical example is epoxy resin reinforced with 60 per cent (volume) of continuous carbon fibres. In refining these materials for optimum properties, much attention has been given to the arrangement of the fibres and to the bonding at the fibre-matrix interface. Fibre diameter has never really been a variable as most commercial fibres are about 10 µm wide. Much smaller diameters have not been available and much larger sizes start to become too coarse. But recent developments¹⁻³ of molecular-scale tubes of carbon and silica point to the possibility of a new class of fine-scale composites.

The trick which accounts for the unusual strength of composite materials lies in two factors. The breaking strength of brittle materials decreases as the size of the largest internal flaw increases. A 10-µm fibre cannot have a transverse flaw bigger than 10 µm, so glass fibres, for example, break at a much higher stress (load divided by cross-sectional

area) than does sheet glass, which has many surface cracks and small internal bubbles. In addition to the fibres' intrinsic strength, any crack in the matrix is deflected by a relatively weak fibre-matrix interface. As long as the sample is in tension parallel to the fibres, the deflected crack becomes harmless and the energy from any impact is absorbed in debonding of the fibres and in the frictional effects as they pull out. It is an example of "united we fall but divided we stand". It seems natural, then, that very small fibres should be good.

Tubular fullerenes have been described recently in this journal¹. These carbon tubes have a diameter of a few nanometres and a length up to several micrometres. Harrington, Goodwin and Kenney^{2,3} have developed methods to produce tubular silicates (see figure) from a rare mineral, litdonite (NaKCuSi₄O₁₀), found only in the crater of Mount Vesuvius. The structure is a tubulation of concatenated 8- and 16-membered rings.

They have now shown that a synthetic litdonite (K₂CuSi₄O₁₀) can be made by forming a glass of the right composition



Tangle of silicate fibres. The smallest (arrowed) have diameters of less than 10 nm and may be individual molecular tubules. (Courtesy of Bruce A. Harrington.)

and devitrifying it. This results in elongated crystals which easily break parallel to the tube axis. The tubes can be separated by replacing the SiO^- groups on the sides of the tubes with $\text{SiOSi}(\text{CH}_3)_3$ groups so that there is only weak van der Waals bonding between them. This can be achieved by slow reaction on the solid silicate. The result is a material which seems to be bundles of fibres with a length of up to 60 μm . The diameter of the bundles ranges up to 800 nm but some fibrils have diameters of only 4 nm, comparable with that expected for a single tube (see figure).

On the nanometre scale, it is not obvious whether to treat these materials as polymers or fibres. There has been an effort over the last few years to make 'molecular composites'. A number of stiff, high-melting-point polymers were developed, principally at Wright Patterson Air Force Materials Laboratories⁴. These had a high strength parallel to the axis of the spun fibre but turned out to be very prone to fibrillation (like the lengthwise 'splitting' of the ends of human hair). This led to the idea that the stiff molecules might be given more lateral strength if they were dispersed into a tougher polymer matrix as if they were fibres in a composite⁵. In practice it proved impossible to get anything like molecular-scale dispersion and nothing has yet come of this approach.

There was also a time when it was hoped that thin (0.1 μm) whiskers of carbon, silicon carbide or metals could be used as reinforcements for composites. These did not need to be continuous if the axial ratio (length/diameter) was large, greater than 20. This approach never really succeeded because the whiskers never became available in sufficient quantities.

Composite theory lays much emphasis on the axial ratio but the effect of fibre

diameter has attracted little interest. At constant axial ratio, or for long fibres, there is no reason to expect a change in strength or stiffness of the composite with fibre diameter. This is provided that the properties of the fibres themselves do not change. Simple models for the pull-out energy and debonding energy of a single fibre scale these as radius cubed, r^3 . Because the number of fibres in a cross-section increases with decreasing fibre diameter only as $1/r^2$, one would expect the work of fracture to decrease as the fibre size is decreased⁶.

Contrary to this, our intuition would say that smaller fibres should be better, at least on the basis of maximum flaw size as outlined above. There should be some optimum strength before molecular scales are reached, because we do not expect a molecular mixture to act like a true composite. The opening up of new routes to inorganic fibres suggest that we should soon be able to put these ideas to the test.

There is a great deal at stake in this issue, as there would be extensive use of such composites to replace metal sheet in cars and elsewhere. A serious problem is the high price of forming composites layer by layer. Small fibres could allow readily mouldable composites to be made, with a consequent dramatic reduction in production costs. □

Paul Calvert is at the Arizona Materials Laboratories, University of Arizona, Tucson, Arizona, USA.

1. Iijima, S. *Nature* **354**, 56–58 (1991).
2. Harrington, B. A. & Kenney, M. E. *Colloids and Surfaces* (in the press).
3. Goodwin, G. B. & Kenney, M. E. *Adv. Chem. Ser.* **224**, 251–263 (1990).
4. Chow, A. W., Bitler, S. P., Penwell, P. E., Osborne, D. J. & Wolfe, J. F. *Macromolecules* **22**, 3514–3520 (1989).
5. Krause, S. J. & Hwang, W.-F. *Mat. Res. Soc. Symp.* **171**, 131–139 (1990).
6. Hull, D. *An Introduction to Composite Materials* (Cambridge University Press, 1981).

Plastic nightmare

THE modern craze for recycling is driven more by the middle-class desire to feel virtuous than by any economic logic. Plastics pose the hardest problem, because the most earnest recycler cannot distinguish one plastic from another. Only homogeneous polymers can be reprocessed at all usefully; the ghastly mixture delivered by most recycling programmes is just about valueless. And yet all the thermoplastics share a similar molecular architecture. Each is made by linking together many copies of a basic monomer unit. The result is a carbon backbone decorated with specific side-chains: phenyl groups for polystyrene, methyl groups for polypropylene, chlorine atoms for PVC and so on. Polyesters and polyamides have phenyl groups and oxygen or nitrogen atoms inserted at intervals into the backbone. A mixture of such similar molecules, says Daedalus, should be recycled on the monomer level. The result would be a homogeneous long-chain polymer, made of all the different monomer units linked together at random.

The chemistry should be fairly simple. Some plastics revert to their original monomer on simple heating. The others could be encouraged to do so by adding the catalyst which originally polymerized them — a catalyst accelerates its back reaction just as much as the forward one. So a mixture of plastics, cunningly catalysed and carefully heated, should slowly homogenize. The various chain-breaking and re-linking reactions should all run simultaneously, scrambling the different monomer units randomly together into long, patternless chains. The mixture would slowly 'anneal' to the thermodynamically most stable copolymer that could be made from the original components.

The resulting chemist's nightmare, Daedalus's 'Massplastic', will be a stable, homogeneous polymer with very consistent physical properties. Simply on statistical grounds, most initial mixtures of recycled plastics should give very similar Massplastics.

Plastics moulders, fabricators and extruders will be able to make just about anything from Massplastic. It will be the perfect average of all they are used to, the ultimate mediocore plastic. As more and more Massplastic products hit the market, it will turn up ever more frequently in recycled material, which will scramble to an even more predictable product. Ultimately the vast bulk of plastic items will be made from this one composition, endlessly recycled. Only its colour will be against it. Averaged and re-averaged from many initial tints, it will inevitably converge to a depressing muddy brown.

David Jones

Introns in sequence tags

SIR — The publication of 'expressed sequence tags' (ESTs) derived from randomly selected clones from commercially obtained brain complementary DNA libraries^{1,2} has received much attention, in part because of its promise to help define hitherto unknown functions encoded by the genome, but also because the genes so defined are now the subject of a patent application by the National Institutes of Health. Indeed, this decision has forced the hand of other agencies in charge of similar projects elsewhere³, not to mention the continuing doubts about the patentability of such sequences. We have now discovered standard cloning artefacts in some of the published ESTs by examining a limited set, suggesting there may be many more.

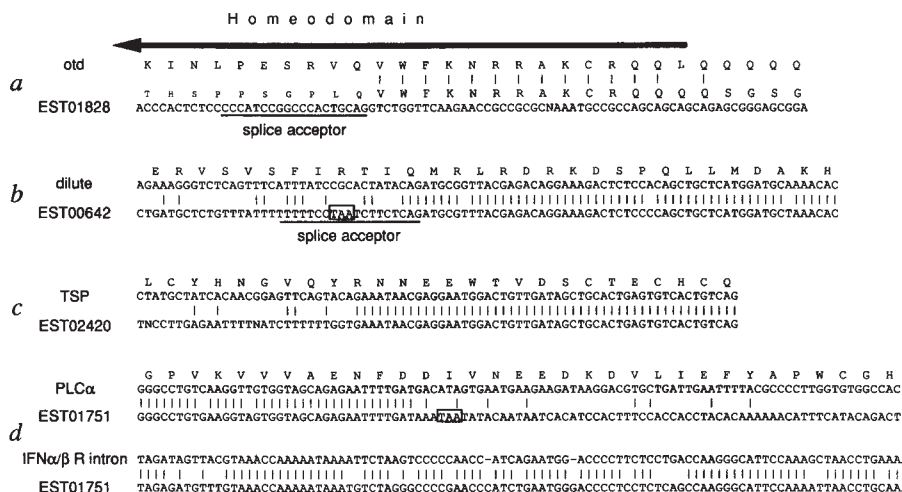
Initially, we examined one EST of interest to us (EST01828, similar to the *Drosophila* homeobox gene *otd*; ref.4), and found that it appears to contain an intron. Subsequently, it seemed that

there were several ESTs reported by Adams *et al.*² that showed a high degree of similarity to known genes (>90%) but over a region much smaller than the mean sequence length of an EST, suggesting abrupt interruptions of the similarity. Of seven such ESTs selected by visual inspection, we found that three seem to contain cloning artefacts (see figure). One terminates its similarity at what appears to be an intron (splice acceptor) and two have apparently unrelated sequences (probably noncoding) joined to a smaller segment containing the region of similarity (see figure). One of these unrelated sequences seems to be a member of a repetitive sequence family. Adams *et al.*² state that they found no evidence that genomic DNA or unspliced cDNAs were major contaminants of their commercially obtained libraries; but given the data presented here we wonder how many entries in the full list are in fact genomic, unspliced or other-

wise corrupted. It would be difficult to estimate this without exhaustive analysis, especially for entries with no sequence similarity, and it would not be an easy problem to solve using computer analysis alone. At least the presence of a poly(A) tail could establish whether a clone is transcribed.

Clones bearing introns may be either cDNAs from unspliced mRNA precursors, or genomic DNA. Certainly, cDNA libraries often contain genomic DNA clones as contaminants. There is thus a finite probability that any of the ESTs are not expressed at all (let alone in the brain). Perhaps, therefore, patent applications and the 'EST' label should be reserved for clones that have been better characterized.

Thomas R. Bürglin
Thomas M. Barnes
 Department of Molecular
 Biology,
 Massachusetts General Hospital and
 Department of Genetics,
 Harvard Medical School,
 Wellman 8, MGH, Boston,
 Massachusetts 02114, USA



Comparison of ESTs with their reported² homologues. *a*, EST01828 compared with *Drosophila otd* (ref. 4). *b*, EST00642 compared with mouse *dilute* (accession no. X57377). *c*, EST02420 compared with human thrombospondin (accession no. M14326). *d*, EST01751 compared with mouse phospholipase C α (accession no. M73329) and a repeated sequence in the first intron of the human interferon α/β receptor gene (accession no. X60459).

a, The sequence of EST01828 with its conceptual translation in the region of similarity to *otd* is shown. Of all homeodomains in the database, EST01828 best matches *otd*. The similarity drops off markedly upstream of the VWF motif, and the conceptual translation product is in small capitals (thick bar, *otd* homeodomain). A consensus splice acceptor is present at the region of divergence. Other members of the *prd*-like class of homeodomains, for example *unc-4* (ref. 5) and *ceh-10* (ref. 6), have an intron in the same position. The sequence of EST01828 further upstream has stop codons and proline residues in all three reading frames in the region where helix 2 might be expected (not shown), which is incompatible with the structure of the homeodomain. *b*, The 286 bp of the EST show similarity to *dilute* over the last 86 bp (reverse orientation). The genomic structure of the mouse *dilute* gene in this region is not known, but EST00642 appears to be the human homologue with an unspliced intron or a genomic clone: a consensus splice acceptor occurs just at the point of sequence divergence, and the upstream sequence contains an in-frame stop codon (boxed). The putative intron has no significant similarity to any other DNA or protein sequence. *c*, The 333 bp of the EST show identity to human thrombospondin over the last 83 bp, except for one frameshift (probably a sequencing error). Before this, the sequences diverge at what does not seem to be a splice site. The nonidentical sequence is 70% (A+T), while the region of identity is close to 50%. It is thus probably noncoding, and the EST probably represents a co-ligation artefact of the cDNA library. The nonidentical sequence has no significant similarity to any other DNA or protein sequence. *d*, The 380 bp EST shows similarity to two sequences. The reported similarity to PLC α occurs over the first 126 bp of the EST. Again, the sequence diverges at what does not appear to be a splice junction. The remainder of the sequence has approximately 70% (A+T) content. Part of this sequence (the last 100 bp) is clearly a member of a human repetitive sequence element, found twice in the first intron of the interferon α/β receptor gene. Because it is only 79% identical, it does not originate from this locus, but presumably from some other dispersed location. Thus the clone is probably a co-ligation artefact. The sequence between these two similarities in the EST shows no further similarity to any DNA or protein sequence.

ADAMS ET AL. REPLY — The quality of a cDNA library is usually assessed pragmatically by whether it contains a particular clone of interest. As a result, incompletely spliced cDNAs and mitochondrial, ribosomal and polyadenylate clones are generally not a hindrance to use of a library. However, any cDNA library constructed from RNA isolated from whole cells, as opposed to the cytoplasm alone, will contain some cDNA to incompletely spliced mRNA, as well as mitochondrial and ribosomal transcripts. Library contamination by genomic DNA is rare, provided sufficient DNase treatment is included in the initial RNA purification step.

In general, a cDNA clone can be conclusively identified as either incompletely spliced or chimaeric only by complete sequencing and direct comparison to mRNA by hybridization, S1 nuclease protection or PCR, together with characterization of all expressed splicing patterns. Sequences that match the consensus patterns for splice sites are very common in exon, intron and intergenic DNA¹; their presence in a cDNA does not provide strong evidence that it is incompletely spliced. While intron positions are conserved in some genes among mammals, they are not conserved universally. We have searched our EST data for known intron sequences, and

POTENTIAL CHIMAERIC, ALTERNATIVE OR UNSPLICED EST SEQUENCES

EST number	Library	DB match
EST01964	FB	TAPA-1 26K cell-surface protein
EST02160	FB	Y-box binding protein YB-1
EST02556	FB	Histone H1(O) + plasma membrane Ca ²⁺ pump chimaera
EST02585	FB	Myoblast cell surface antigen 24.1D5
EST01453	FB	γ -actin + 28 SrRNA
EST00362	HC	ret protooncogene + mitochondrial cytochrome B
EST00949	HC	Vesicle-associated membrane protein-2
EST00971	HC	G ₀ α
EST01338	HC	NF- κ β
EST01785	HC	Retinaldehyde binding protein (EcoRI linker)

FB, fetal brain library; HC, hippocampus library, both from Stratagene. DB match are the GenBank match(es) for the particular EST clone.

have examined ESTs that exactly match human genes for which genomic sequences are available for the presence of introns; neither of these methods yielded evidence for unspliced mRNAs. As we have noted^{2,3}, the libraries we have used contain substantial numbers of interspersed repetitive sequences, which may be from untranslated 3' tails of mRNAs, unspliced introns or fortuitous transcripts from nearby Pol II promoters.

We have observed several instances of database matches that appear to end prematurely. These sequences may represent chimaeric clones, unspliced introns or mRNAs that are alternatively spliced compared to the sequence in GenBank. One technical limitation which promotes overprediction of premature overlap termination is the use of BLAST as a database search tool. BLAST is unable to extend matches through frameshifts or gaps in a sequence. Therefore, only one of possibly several BLAST matches is reported in the tables in refs 2,3. In addition to those noted by Bürglin and Barnes, the ESTs listed in the table below appear to differ in structure from the mRNA sequences reported in GenBank.

We have analysed more than 1,000 of the ESTs by restriction digestion to determine insert size and identify clones that contain internal *EcoRI* sites, and hence may be chimaeras. A substantial fraction (30–40%) of the clones from the hippocampus library appear to be chimaeric, and we have discontinued sequencing from this library. We are working both at NIH and in collaboration with other laboratories to develop high-quality cDNA libraries from several tissues specifically for EST sequencing. Advances in the state of the art in cDNA library construction should help to eliminate difficulties in interpretation of short database matches and facilitate immediate full-length cDNA cloning.

The restriction digestion data and other data that we did not report because of space limitations are included in our EST database, which is freely available to all researchers by anonymous file

transfer (ftp) from briggs.ninds.nih.gov. Several dozen laboratories throughout North America, Europe, Japan and Australia have already obtained copies of this complete database, which we continually update with new information on the clones and sequences. We welcome any information regarding ESTs from users, and will include this information in the database as appropriate.

Each of the ESTs that Bürglin and Barnes mention contains sufficient coding sequence (as shown by sequence alignments) unambiguously to identify a human gene, and hence to permit rapid isolation, complete sequencing and functional characterization of that gene. Providing such identifying sequence information to the biological and medical communities quickly and inexpensively is one primary purpose of our EST sequencing project.

The critical and defining force behind the NIH patent application on cDNA sequences was a concern about the future patentability of the full-length cDNA clones and of products derived from cDNA clones and sequences, rather than about patentability of unknown DNA fragments. The current state of the art in molecular biology makes cloning and sequencing a complete gene standard practice — hence “obvious” — given a short part of that sequence. Because patent law protects “obvious” extensions, ESTs in the public domain could interfere with future patent claims for diagnostic, therapeutic or forensic applications of the sequences. Resolutions of this issue, and of the general issue of the most effective means of encouraging beneficial applications from the information generated by the genome project, is a matter of paramount concern. For example, how can government encourage the development of therapeutics for the 5,000 single-gene-based genetic disorders when each dis-

ease appears only in a few thousand individuals? Participation and informed discussion⁴ by the scientific community is clearly appropriate.

Mark D. Adams

Chris Fields

J. Craig Venter

Section of Receptor Biochemistry and Molecular Biology,
National Institute of Neurological Disorders and Stroke,
National Institutes of Health,
Park Building, Room 405,
Bethesda, Maryland 20892, USA

Misrepresenting paradise

SIR — I am the unidentified ‘geologist’ accused of despoiling parts of Ellesmere Island with “the largest collection of garbage” that Dr France encountered on a ski-traverse¹. France quotes one of my papers concerning conservation of the high Arctic and implies that my intention was selectively to discourage tourism because of the area’s exceptional environmental sensitivity. On both issues France has misled *Nature*’s readership.

First, the ‘garbage’ to which France alludes is a former cache that was left in the field secured in metal and wooden boxes in 1976. Limited financial support precluded an extra aircraft flight that summer to remove it and our plan was to reuse this equipment. The cache was not abandoned, and what was garbage to France was sustenance to me! Until 1984, colleagues of mine worked at this site and used my freight canoe. The cache was revisited as recently as 1988 and was still secure, but the boat had disappeared, possibly stolen. In any event, the cache was removed to Resolute Bay more than a year ago, so I fail to see the point of France’s argument.

My articles about the preservation of the high Arctic environment never implied that tourists should be excluded while scientists enjoyed the area as a private sanctuary. I have worked on Ellesmere Island for more than 20 years and, together with my graduate students, we have had several hundred camps. All the items we have taken in have gone out, down to the last pieces of string.

Finally, France unwittingly reveals his unfamiliarity with Arctic research when he says that the Polar Continental Shelf Project should have a ‘garbage in garbage out’ policy. It has been doing so for more than a decade. France’s naivety about Arctic biology is similarly revealed in his statement that the land has only “recently emerged from its protective cradle of ice”. This ignores all the research that has demonstrated that large ice-free areas served as biological refugia

- Harris, N. & Senapathy, P. *Nucleic Acids Res.* **18**, 3015–3019 (1990).
- Adams, M. et al. *Science* **252**, 1651–1656 (1991).
- Adams, M. et al. *Nature* **355**, 632–634 (1992).
- Gillis, A. M. *BioScience* **42**, 336–339 (1992).

for tens of thousands of years²⁻⁵.

Most fieldwork in the Arctic is conducted by conscientious researchers who cherish this environment. It is careless and misleading to imply that the Arctic is 'riddled' with their garbage which serves as a testimony to their 'arrogance'. That France encountered only 208 objects across 1,200 km (six of them being cigarette butts) affirms that he was in paradise but simply failed to recognize it.

John England

Department of Geography,
University of Alberta,
Edmonton, Alberta,
Canada T6G 2H4

1. France, R. *Nature* **355**, 504 (1992).
2. England, J. & Bradley, R. S. *Science* **200**, 265-270 (1978).
3. England, J. *Can. Geog.* **103**(3) 8-17 (1983).
4. Stewart, T. G. & England, J. *Boreas* **15**, 25-31 (1986).
5. Lemmen, D. S. *Can. J. Earth Sci.* **26**, 2578-2590 (1989).

Mercury and gold pollution

SIR — Nriagu *et al.*¹ report mercury pollution in the Madeira river in the Brazilian Amazon stemming from the gold rush, mercury being used to extract the gold by amalgamation. The same method of gold extraction was used in mining operations in the last century and the early part of this century in the Dolgellau gold belt of Gwynedd, north Wales in the United Kingdom. Although the scale of the operation in Wales was much smaller, with about 3.7 tonnes of gold being produced between 1860 and 1916 (ref. 2), it has resulted in enhanced levels of mercury in sediments of the Mawddach river, which drains the gold belt area.

We collected sediment samples from the bed of the Mawddach river near one of the most productive mines of the area, Gwynfynedd. Samples were collected during a period of low flow in January 1991 and air-dried. The sediment was sieved and divided into seven

fractions, the mercury being concentrated in the fractions between 250 and 63 μm , the highest concentrations being in the 180-125- μm fraction. Here we report an analysis of the 250-180 and the 180-125 μm fractions.

From the data presented in the table, it is apparent that although gold is concentrated in the coarser of the two fractions, mercury is mainly concentrated in the finer of the fractions. There is little correlation between levels of the two metals in either fraction, which would tend to suggest that they occur independently and not in amalgam form.

At the Gwynfynedd operation the processing plant was located at the confluence of the Afon Gain with the Mawddach river, about 600 m from the mine site. Our data show that the highest levels of both mercury and gold occur in the sediments from the Mawddach and the Gain at and just below this point, concentrations showing a marked decline with distance downstream. The relatively enhanced levels of mercury and gold which occur near the mine site probably reflect contamination from the mine. Considerable amounts of sphalerite occur in Gwynfynedd mine, and this mineral contains appreciable amounts of mercury³.

Although much of the mercury and gold in the sediments probably results from release during extraction, the extremely high values found for both elements in tailings waste near the old mill site suggest that this may be a continuing source of contamination in the river sediments. These tailings are fine and do not support any vegetation and are therefore relatively easily eroded. We found enhanced levels of both mercury and gold in a small channel draining from the tailings into the Mawddach, but found no evidence of enhanced mercury contents of river water.

There is much concern about mercury pollution of river systems resulting from gold extraction in areas such as Brazil and Indonesia. The elevated levels of mercury in stream sediments in this re-

gion of north Wales where gold mining was on a fairly small scale more than 80 years ago suggests that this contaminant metal has a fairly long residence time in river sediments. Any disturbance of these sediments even many years after the cessation of gold extraction may result in environmental problems.

Ronald Fuge

Nicholas J. G. Pearce

William T. Perkins

*Geochemistry and Hydrology
Research Group,
Institute of Earth Studies,
University of Wales,
Aberystwyth,
Dyfed SY23 3DB, UK*

1. Nriagu, J. O., Pfeiffer, W. C., Malm, O., Souza, C. M. M. & Mierle, G. *Nature* **356**, 389 (1992).
2. Hall, G. W. *The Gold Mines of Merioneth 2nd edn* (Griffin, Kingston, 1986).
3. Levinson, A. A. *Introduction to Exploration Geochemistry* (Applied Publishing, Wilmette, Illinois, 1980).

Alpha-particles and leukaemia

SIR — The interesting demonstration of nonclonal transmitted genetic instability in haematopoietic stem cells after α -particle irradiation¹ was linked with the possibility that it might increase the expected induction of leukaemia by low levels of α -particle irradiation^{1,2} in circumstances when most of the exposed cells escape direct radiation damage. It was not proposed that the transmitted instability occurred in particular chromosomes but that the instability could, on occasion, cause a specific leukaemia-initiating lesion and so augment the frequency of induced leukaemia.

But the experimental observations could imply the converse expectation. An induced genetic instability of cellular chromosomes in general, maintained over many cell generations after exposure to α -particles, would be expected to affect the progeny not only of those cells directly damaged at the time of exposure but also of cells in which specific leukaemia-initiating lesions had been induced at that time. Nonclonal chromosomal instability in an initiated cell, especially if causing multiple aberrations (Fig. 2 of ref. 1), leading to its premature death, would decrease the expectation of overt leukaemia.

Observations on human subjects after α -particle irradiation of haematopoietic tissues show that induced leukaemia is uncommon³⁻⁵, much less frequent than orthodox radiobiological concepts would suggest. This is so for dial painters contaminated with ²²⁶Ra and ²²⁸Ra and for children and adults receiving multiple injections of ²²⁴Ra (refs 6,7), all groups with high incidences of induced bone sarcoma, and in other groups with expo-

MERCURY AND GOLD CONCENTRATIONS IN SEDIMENTS FROM THE GWYNFYNNEDD AREA

Sample	125-180 μm		180-250 μm	
	Au	Hg	Au	Hg
Mawddach 1.2 km above Gwynfynedd	ND	ND	0.14	0.09
Mawddach immediately below mine	0.21	0.13	ND	0.12
Mawddach 400 m below mine	0.11	0.19	0.10	0.15
Mawddach at confluence with Gain	0.45	3.68	0.18	1.26
Gain at confluence with Mawddach	ND	1.15	0.65	2.70
Mawddach 40 m below confluence	5.10	1.60	23.8	0.79
Mawddach 350 m below confluence	ND	0.14	0.06	0.11
Mawddach 2.4 km below confluence	0.12	0.13	0.10	0.05
Mawddach 4.2 km below confluence	—	—	ND	0.05
Tailings materials at old mill site	15.1	6.07	44.0	6.03
Small drainage channel below mill	0.20	0.46	0.26	0.27

ND, not detected (detection limit 0.05 mg kg⁻¹). Dashes, not determined. Values are in mg kg⁻¹.

tures too low or too recent for bone sarcomas to be seen^{8,9}.

I was once asked to discuss why leukaemia was not found after α -particle irradiation. A re-examination of the data provided strong circumstantial evidence that unrecognized leukaemia had occurred in dial painters but only in those few subjects with the highest body burdens¹⁰, not in the several hundred dial painters in the United States and United Kingdom with body burdens 100–1,000 times less than the highest. Conversely leukaemia, but not bone sarcoma, has often been seen after injection of Thorotrast³ (a preparation of colloidal ²³²Th), but dose in individual cells within 100 μ m of a Thorotrast deposit is rarely from a single α -particle, that is, rarely as low as 1 Gy. Each deposit emits a series of α -particles as ²³²Th and its daughters undergo a sequence of nuclear transformations.

When experimental α -particle irradiation gives a linear dose-response without threshold and low LET, for example, X-ray irradiation gives a curvilinear dose-response without threshold, the ratio of dose for equal effect, the RBE, must increase asymptotically to an infinite value as dose tends to zero (compare the figure in ref. 2). Such contrasting dose-responses are often found, but the legitimacy of extrapolation below the lowest dose used in an experiment must depend on the presumed mechanism underlying the particular radiation effect under consideration (about which there may be little agreement). Recent work on cell survival at very high levels (after fractionated X-ray doses < 0.5 Gy in tissues *in vivo*¹¹ and after lower single doses in V79 cells in culture (B. Marples and M. C. Joiner, personal communication)) shows that below a critical dose level responses to low LET radiation can change abruptly with a corresponding progressive decrease in RBE for neutrons, so denying both legitimacy of extrapolation from a dose-response when its lowest dose is not sufficiently low and the inference that RBE must increase monotonically as damage (dose) decreases (see ref. 2).

More generally, when considering leukaemogenesis and carcinogenesis, it seems more acceptable to base inferences on the actual occurrence *in vivo* of leukaemia (or cancer as the case may be) on evidence on native cells in tissues, rather than on extrapolation from evidence on transformed or untransformed cells maintained in culture. Nevertheless the new and unexpected findings on genetic instability after *in vitro* exposure to α -particles¹ do have a special interest as a possible explanation of an unexpectedly low incidence of leukaemia after low-level exposure of humans to α -particles.

The proportion of clones without aberrations after 0.25, 0.5 and 1 Gy was 0.6, 0.5 and 0.6, respectively¹ (no dose response). It is not known whether the observed chromosomal instability in a clonogenic cell was caused by the action of a single α -particle or more than one α -particle.

R. H. Mole*

Bayworth Lane,

Boars Hill, Oxford OX1 5DF, UK

* Sadly, Dr Mole died a few days after submitting this letter.

1. Kadhim, M. A. et al. *Nature* **355**, 738–740 (1992).
2. Evans, H. J. *Nature* **355**, 674–675 (1992).
3. Mole, R. H. in *The Radiobiology of Radium and Thorotrast* (eds Gossner, W., Gerber, G. B., Hagen, U. & Luz, A.) 1–13 (Urban & Schwarzenberg, Munich, 1986).
4. Taylor, D. M., Mays, C. M., Gerber, G. B. & Thomas, R. G. (eds) *Risks from Radium and Thorotrast Report No. 21* (British Institute of Radiology, London, 1989).
5. Priest, N. D. in *Risks from Radium and Thorotrast* (eds Taylor, D. M. et al.) 159–165 (British Institute of Radiology, London, 1989).
6. Spiess, H., Mays, C. W. & Chmelevsky, D. in *Risks from Radium and Thorotrast* (eds Taylor, D. M. et al.) 7–12 (British Institute of Radiology, London, 1989).
7. Stebbings, J. H., Lucas, H. F. & Stehney, A. F. in *Epidemiology Applied to Health Physics Proc. 16th Topical Meeting Health Phys. Soc. (Albuquerque, New Mexico, CONF-830101)*.
8. Wick, R. R. & Gossner, W. in *Risks from Radium and Thorotrast* (eds Taylor, D. M. et al.) 25–28 (British Institute of Radiology, London, 1989).
9. Baverstock, K. & Papworth, D. W. in *Risks from Radium and Thorotrast* (eds Taylor, D. M. et al.) 72–76 (British Institute of Radiology, London, 1989).
10. Mole, R. H. in *Risks from Radium and Thorotrast* (eds Taylor, D. M. et al.) 153–158 (British Institute of Radiology, London, 1989).
11. Joiner, M. C. in *The Scientific Basis for Modern Radiotherapy* (ed. McNally, N. J.) BIR Report No. 19, 20–6 (British Institute of Radiology, London, 1989).

SIR — The results of Kadhim *et al.*¹ on the incidence of nonclonal chromosomal aberrations in stem cells of the bone marrow subject to two or less α -particle passages on the average, but not in cells exposed to γ -rays, and the interpretation of radiation-induced chromosomal instability are of great potential importance in a wider context than leukaemogenesis. A similar interpretation can be advanced to explain some earlier results obtained by Luning and colleagues^{2,3} in studies on the induction of dominant lethality in mice by ²³⁹Pu α -particles.

In these, male mice were injected with ²³⁹Pu citrate solution, sequentially mated to unirradiated females every week and dominant lethality (expressed as early and late intrauterine death) was measured at 18 days post-mating. Late as well as early deaths were observed in conceptions deriving from irradiated post-meiotic and peri-meiotic germ cells.

There were two surprising findings: (1) in successive mating intervals, there was a shift in favour of late deaths, in contrast to the situation recorded for X-irradiation (a shift in favour of early deaths⁴); and (2) when females were allowed to litter and the F₁ males from matings in the 9th, 14th and 16th weeks (those descended from irradiated stem cell spermatogonia) were used in dominant lethal tests, there were significant increases in both late and early deaths,

suggesting 'transmission' of dominant lethality. None of the 410 F₁ males (and 122 controls) showed evidence of semi-sterility, thus excluding radiation-induced reciprocal translocations as a causal factor. In common with the findings of Kadhim *et al.*¹, such 'transmission of dominant lethality' was not observed with X-rays in the same strain of mice.

Nonclonal aberrations cannot be explained in terms of stable sequence changes in the genomic DNA. However, a mechanism can be envisaged on the basis of the concept of the existence of a dynamic steady state between DNA degradation and repair as it applies to radiobiology⁵. Whether, in estimating genetic risks of α -particle exposures, one should take into account the possibility of genomic instability in addition to RBE considerations, and if so, how, can only be answered by further studies.

K. F. Baverstock

WHO European Centre for Environment and Health,

Rome, Italy

K. Sankaranarayanan

Department of Radiation Genetics and Chemical Mutagenesis, Sylvius Laboratory, University of Leiden, Leiden, The Netherlands

1. Kadhim, M. A. et al. *Nature* **355**, 738–740 (1992).
2. Luning, K. G., Frölen, H. & Nilsson, A. in *Biological and Environmental Effects of Low Level Radiation* **1**, 33–49 (IAEA, Vienna, 1976).
3. Luning, K. G., Frölen, H. & Nilsson, A. *Mut. Res.* **34**, 539–542 (1976).
4. Bateman, A. J. *Heredity* **12**, 213–232 (1958).
5. Baverstock, K. F. *Int. J. Radiat. Biol.* **60**, 581–595 (1991).

WRIGHT REPLIES — It is quite likely that many of the chromosome aberrations we reported¹ may represent lethal events for the murine cells, and Robin Mole's proposition that this is a possible explanation of an unexpectedly low incidence of leukaemia after low-level exposure of humans to α -particles is one of several interesting speculative implications of our data.

Other investigations in the MRC Radiobiology Unit using the strain of mice from which we obtained the marrow cells have demonstrated that injection of bone-seeking α -emitters resulted in the development of acute myeloid leukaemia in some of the animals². It is possible that the phenomenon of chromosomal instability contributes to the development of leukaemia in such animals, but it should be noted that our experiments do not address the question of leukaemia incidence.

Eric Wright

MRC Radiobiology Unit, Chilton, Didcot, Oxford OX11 0RD, UK

1. Kadhim, M. A. et al. *Nature* **355**, 738–740 (1992).
2. Humphreys, E. R. et al. *Int. J. Radiat. Biol.* **51**, 331–339 (1987).

The environment as a commodity

Wilfred Beckerman

This week's "Earth Summit" in Rio has precipitated the publication of many books exploring the link between the environment and the economy. Here is an economist's assessment of their use to scientists.

TWENTY years ago I gave an inaugural lecture at University College London entitled "Economists, scientists and environmental disaster". It was a peak year for 'public concern' about the alleged conflict between economic growth and environmental protection. More specifically, this unease was restricted to those in the more advanced countries of the world, or rather to the more affluent inhabitants there.

The environmental lobby has since become increasingly active and influential. For as societies become richer, concern about the environment naturally and rightly moves up in their hierarchy of objectives. More recently, there has also been a mounting fear of global warming and of damage to the ozone layer. Whether or not these worries are fully justified, they have helped spawn an enormous number of books and pamphlets prophesying impending environmental catastrophe unless the human race takes drastic action: we must renounce our wicked desire to improve our material prosperity, or at least cut in half emissions of carbon dioxide and hand over vast amounts of money to agencies devoted to protecting us from our folly.

It is true, of course, that our environment is being polluted on a vaster scale than ever before and at an increasingly rapid rate, as is well set out by Jim MacNeill *et al.* in *Beyond Interdependence*. So we are right to be concerned. It is also true that all the primary data on, and analysis of, the problem come

from scientists, whose contributions to the debate — particularly in immensely complex areas such as climate change — represent an intellectual achievement of the highest calibre. But that does not mean that scientists have special insight into society's choices or moral obligations in these areas. Steps to prevent global warming might well ensure that the economic welfare of people living at the end of the next century is, say, 4.4 times as high as it is now, instead of 4.2 times had no action been taken. But prevention could also impose heavy costs on current generations, particularly on poorer people in developing countries who, as the World Bank's *World Development Report 1992* admirably documents, have far more urgent environmental needs. Here, scientists have no decisive say in determining whether prevention is justified.

Nor for that matter do economists, although they, at least, have developed precise criteria for deciding how best to distribute limited resources, fitting alternative courses of action into the economist's framework of balancing costs and benefits at the margin. And they have long recognized that the free market does not work perfectly and that among the imperfections are 'externalities', of which environmental pollution is a glaring example. So it is inevitable that in the pile of publications carried along on the environmental bandwagon, there are several books that purport to provide a guide to the contribution that econo-

mics can make to the debate.

Some of these publications may be reasonably useful to the non-economist, but suffer from having been produced too hastily. For example, one's confidence in M. D. Young's mastery of the arithmetic he uses in *Sustainable Investment and Resource Use* is weakened by his telling us, on page 12, that "sustainable development has become a household word". Sloppiness also abounds in the generally useful *Economics of Natural Resources and the Environment* by E. Kula, which succeeds in misspelling several people's names, including mine and the editor of *Nature*'s. But both books are helpful in at least giving non-economists a fairly balanced picture of what environmental economics is about.

Unconventional economics

The same cannot be said of *Ecological Economics* edited by R. Costanza. The few professional economists who contribute to this volume are totally unrepresentative of the many now working on environmental issues. For example, Kenneth Building's and Herman Daly's dire predictions of what will happen to the environment in the absence of drastic changes in attitudes and policies are by no means shared by most economists, and nowhere in the volume is the alternative point of view fairly represented.

This applies also to R. B. Norgaard and R. B. Howarth's discussion in the same volume of the relationship between "sustainability and discounting the future". Although these authors do raise valid questions about the ethical value judgements underlying the economist's convention of discounting future costs and benefits in order to arrive at present values, they reach conclusions that would be disputed by the overwhelming majority of professional economists. The authors recognize that some distinguished economists such as J. Stiglitz and the Nobel laureate Robert Solow, have a different view about the discount rate, but simply to say that "it is difficult to understand why academic economists have put so much emphasis on discounting" is ridiculous. There is no difficulty at all. Almost all economists believe that it makes sense to discount the future given (1) uncertainty and (2) the assumption that an extra unit of income provides less 'utility' to a richer person

Books under review

■ **Beyond Interdependence: The Meshing of the World's Economy and the Earth's Ecology.** By Jim MacNeill, Peter Winsemius and Taizo Yakushiji. Oxford University Press: 1992. Pp. 159. £22.50, \$24.95 (hbk); £6.99, \$7.95 (pbk).

■ **World Development Report 1992.** Oxford University Press/World Bank: 1992. Pp. 344. £27.50, \$32.95 (hbk); £12.50, \$16.95 (pbk).

■ **Sustainable Investment and Resource Use: Equity, Environmental Integrity and Economic Efficiency.** By M. D. Young. Parthenon: 1992. Pp. 176. £35, \$65.

■ **Economics of Natural Resources and the Environment.** By E. Kula. Chapman and Hall: 1992. Pp. 287. £40.

■ **Ecological Economics: The Science and Management of Sustainability.**

Edited by Robert Costanza. Columbia University Press: 1991. Pp. 525. \$57.

■ **National Income and Nature: Externalities, Growth and Steady State.** Edited by J. J. Krabbe and W. J. M. Heijman. Kluwer: 1992. Pp. 232. Dfl. 260, £97.50, \$56.

■ **Global Warming: Economic Policy Responses.** Edited by Rudiger Dornbusch and James M. Poterba. MIT Press: 1991. Pp. 389. \$26.95, £29.95.

■ **Beyond the Limits: Confronting Global Collapse, Envisioning a Sustainable Future.** By Donella H. Meadows, Dennis L. Meadows and Jorgen Randers. Chelsea Green/Earthscan: 1992. Pp. 320. \$19.95, £19.95 (hbk); £9.95 (pbk).

than to a poorer person and that future societies will be far richer than those today.

Perhaps one of the most dangerous examples of the unconventional economics in this volume appears in R. Heuting's piece on how to correct national income estimates for environmental damage, which this author also presents in *National Income and Nature*. Heuting, whose *non sequiturs* include assertions such as "Defining production growth as economic growth means defining economics as production" (page 23), generally seems to have great difficulty writing two logical sentences in sequence. Nevertheless, he has had enough influence in his own country, the Netherlands, to persuade its Central Bureau of Statistics to waste scarce resources in a premature attempt to incorporate environmental factors into estimates of gross national product. Almost all economic statisticians would much prefer to concentrate on improving the currently fragile estimates of conventional GNP figures.

Economists are commonly accused of disagreeing among themselves on almost any economic question put to them. As far as the environment is concerned, however, there is probably far more disagreement among scientists, and a surprisingly wide area of agreement among economists. *Ecological Economics* should not be taken by scientists as bearing much relationship to this consensus.

In fact, of the books here, the only one that gives an accurate and scholarly representation of how economists tackle the problem of weighing up the costs and benefits of alternative environmental policies is *Global Warming and Economic Policy Responses*, edited by Rudiger Dornbusch and James Poterba, economists at the Massachusetts Institute of Technology. This volume contains what is probably the best attempt so far (by W. D. Nordhaus) to estimate not merely the economic costs and benefits of a given increase (usually a doubling) in carbon dioxide concentrations, but the whole schedule of costs and benefits for alternative rates of emission in order to arrive at optimum policies.

Benefits of pollution

To the non-economist, it may seem strange to talk about the 'benefit' of carbon dioxide emissions, or of any pollution for that matter. But if production is not to be reduced, then the benefit is the saving that society makes through not having to replace the 'use' of the environment (which is what pollution amounts to) by other factors of production (labour, capital and raw materials). These factors can then be used to produce other goods and services that contribute to human welfare, such

as food, clothing, houses, schools, hospitals, household durable items and transport. Using the environment instead frees them for these purposes and is therefore a benefit. At the same time, the damage done by pollution often imposes a cost on society. It is thus reasonable to allow pollution to use the environment up to the point at which the marginal benefits of further use (that is, further pollution) are just matched by the marginal costs of the damage done.

Of course, to apply this methodology to global warming is immensely complicated. It is difficult to attach values to the damage caused by global warming and to the costs of preventing it; and global warming is a phenomenon that builds up over time, so the optimum degree by which emissions should be reduced is very sensitive to the choice of discount rate. Nordhaus is a master of these complexities though, and he succeeds in outlining various alternative estimates. His main estimate is that it would be desirable to aim at "a total reduction of about 12 per cent of GHG [greenhouse gas] emissions, including a large reduction in CFCs [chlorofluorocarbons] and a small reduction in CO₂ emissions".

This result may well surprise many people. But most expert analysis of the economic damage caused by a doubling of the global carbon dioxide concentration indicates that it will be a negligible proportion of total world output. This is partly because the sector most vulnerable, agriculture, is hardly likely to be affected taking the world as a whole: improvements in the yields of crops in some areas at high latitudes will offset deterioration in others (not to mention the boost in yields caused by increases in concentrations of carbon dioxide). Of course, the effects on individual regions will vary greatly, and when the climate models are able to provide much more precise predictions of regional and local effects, these estimates will have to be revised. But it is impossible to forecast in what direction they will be altered. On the other hand, although it seems likely that unchecked climate change will not produce the catastrophic effects widely proclaimed by the environmental pressure groups, including the rise in sea level and an increase in the frequency of storms, the investment needed to reduce the danger of climate change is also only very small relative to total world output. This is all brilliantly spelled out by Thomas Schelling in *Global Warming*. Most of the other contributions to this volume are equally packed with serious and scholarly analyses of the economics of global warming, and there is no doubt that the book is by far the best buy on the market for anybody who wants some insight into how economists deal with

environmental problems.

By contrast, *Beyond the Limits* is a wonderful example of how to blind the public with pseudoscience, with the aid of lots of diagrams showing the outcomes of alternative computer simulations into which absurd assumptions have been fed — inevitably this leads to absurd results. But what is astonishing about the book is the audacity of its authors, who 20 years ago published a similar apocalyptic volume entitled *The Limits to Growth*. That book contained a variety of doomsday predictions that were not only rightly savaged by critics at the time but some of which have already been shown to be erroneous.

Garbage in, garbage out

For example, to arrive at the conclusion that continued economic growth would sooner or later be constrained by diminishing reserves of natural resources, the authors compared existing reserves, or some finite multiple of them (five times was the upper limit), with the consumption rates. Their computer arrived at the astonishing conclusion that if demand for a given material resource rose exponentially and supply was fixed, demand would eventually exceed supply. Surprise, surprise! The power of science and computerized systems analysis stunned the media and many politicians.

But in the 20 years since the authors made these predictions, cumulative consumption of some of the materials on their list has actually exceeded initial reserves, yet we now have more of the materials than we had at the outset. The latest study does pay some lip service to the need to allow for favourable feedbacks in the form of rising prices of scarcer raw materials; but the authors still conclude that the claims on capital needed to respond to the increasing scarcity of raw materials is such that economic growth will soon be severely constrained. This claim has, of course, been made time and time again over the past 2,000 years or more and has always been proven wrong. Anyway, as a matter of simple logic, if resources are really finite in any meaningful sense, then economizing in their use or slowing economic growth will not save us in the long run.

Perhaps we will learn to adapt, as we shall to global warming. After all, as I said in my lecture 20 years ago, the human race has managed to get along splendidly for well over a century without any supplies of Beckermonium, a product named after my grandfather who failed to discover it more than 100 years ago. □

Wilfred Beckerman is a fellow in economics at Balliol College, University of Oxford, Oxford OX1 3BJ, UK.

Ground rules

N. W. Pirie

Out of the Earth: Civilization and the Life of the Soil. By Daniel Hillel. Aurum/Free Press: 1992. Pp. 321. £19.95, \$22.50.

"WHERE there is no vision, the people perish"; or, as a marginal note in the better bibles puts it, "will become unruly". Many recent events justify the suggested variant, but the first is more relevant in the present context. That biblical exegesis comes naturally because Hillel's book, although not containing that particular quotation, is liberally sprinkled with passages from religious and other myths and allegedly inspired writings. To be a perfect epigraph, 'vision' should be replaced by 'knowledge' or 'use of existing knowledge'. Hillel explains that we now know how to avoid making the old mistakes, such as eroding soil and contaminating water with salts and the products of modern industry. And we have research techniques that should keep us from making new mistakes. But knowledge and research are inadequately exploited. Perhaps it is time we became more unruly.

Although the word 'earth' is heavily capitalized in the title of the book, and 'soil' is in the subtitle, discussion of their nature, origins and compositions occupies only seven pages: water is the main theme. Irrigation and erosion are the subjects that get most citations in the index, and there are nearly as many for groundwater and rainfall. Included is a useful bibliography of about 200 references.

Hillel judges beneficial modification of rainfall to be unlikely. He briefly discusses the possibility that deforestation and overgrazing may diminish rainfall, by increasing albedo so less warm air rises from sunlit bare ground, and rightly dismisses it as unproven. But the suggestion that complete plant cover would persist if established on a meteorologically significant scale in an arid area and if protected from overexploitation, deserves thorough testing. Irrigated perennial leaf crops which, unlike seed crops, do not leave bare ground after harvesting, might later start a cycle of adequate rainfall.

Delaying runoff

Several once populous and fertile regions are now sparsely inhabited. The cause of this change is often obscure: sometimes it was increasing aridity. Climates can change without human mismanagement, but mismanagement causes failures in crop growth without a climatic change. Hillel tells how, as a

Scorched Earth — this year's report from the World Resources Institute estimates average annual deforestation in West Africa to be 36 times faster than reforestation.

preface to giving evidence on soil erosion to a US committee, two experts spread a towel on the table, poured a glass of water onto it, picked up the towel and poured the same amount of water onto the bare table. This convinced the wetted congressmen that soil could, in a similar way, delay runoff and so increase the agricultural efficacy of rain. He says that it is now difficult in the United States to find potentially arable land from which soil has not been lost, because mechanical disturbance and periods when the ground was bare have accelerated erosion. Yet techniques that promote erosion are still extensively used, making dust storms and silted-up ditches common. And even with the availability of modern machinery, the ancient remedies of terracing steep slopes and using contoured ridges to delay runoff from gentle slopes are now considered to be uneconomic.

Hillel gives many examples of this neglect of existing knowledge in a survey of contemporary use of the ancient art of irrigation. Between 1965 and 1990, the total area of irrigated land nearly doubled. But much of the water was used with an efficiency of only 30 per cent, whereas 80 per cent is possible. Between 1950 and 1970, water use in Israel diminished by 25 per cent per hectare; nevertheless, yields increased. Hillel finds much to praise in Israel's water policy, but he is forthright in condemning the draining of the Huleh marshes. He is similarly doubtful about the feasibility and benefits of completing the Jonglei canal, which is expected to diminish evaporation in the Sudd and thus increase the amount of water reaching Egypt. In so far as his arguments for swamp conservation depend on such factors as preservation of wildlife and the supply of thatching material, they are not likely to prove influential with economists. A better argument is that, with

more research, lush swamp vegetation could become a useful source of protein for people and other nonruminant animals.

Preventing salinization

Much of the water used for irrigation is pumped from underground stocks that are not being fully replenished by current rainfall. The argument that it is foolish to depend on water that has accumulated during past centuries is irrefutable. It has a counterpart in the suggestion that some irrigation projects had to be abandoned in antiquity because they depended on meltwater from glaciers left over from the last ice age. Salinization is an inevitable consequence of irrigation wherever there is no outward drainage — salts remain when water evaporates. Even when drainage is possible, those living downstream of the irrigated area — for example, Mexicans relying on the salinized Colorado — may object to this pollution. Ultimately, it may be necessary to sink bore-holes and pump salinized water out to sea.

As is now usual in semipopular science books, the 'greenhouse effect' is discussed. The elementary error of regarding climax rainforest as a source of oxygen and a sink for carbon dioxide is avoided: these processes are precisely reversed by the rotting vegetation on the ground. As Lawes pointed out 140 years ago, although a forest carries more carbon per hectare than a wheat field, it fixes no more carbon dioxide per hour. Hillel accepts an estimate from the Food and Agriculture Organization that shifting cultivation is responsible for 35 to 70 per cent of deforestation and argues that this is acceptable provided the affected plots are given adequate time to regenerate: it is permanent deforestation for timber or firewood that he condemns. In an eco-friendly slash-and-burn system, the burning would go on in the kitchen, with ash

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

returned to the plot. A good way to get rid of carbon dioxide is to use water to build up soil organic matter by farming the desert, and to restrict drainage in some areas so as to make peat — if that can be managed without making methane. It is surely legitimate to wonder whether it will ever be practical to make carbon-based roads and buildings instead of liberating carbon dioxide by making cement.

Hillel has a pleasant scepticism. He suggests that the menace of spreading deserts has been exaggerated, as is the harm done by goats. Goats, like people, create problems when too numerous. But people cause most trouble because of their greed, stupidity and unwillingness to use existing knowledge; goats, by contrast, are difficult to control because of their intelligence. For our convenience, we need a strain of goats that retain the remarkable digestive system of current types, but that are as stupid as sheep. Several other interesting research problems, aimed at giving the world a form more congenial to us, are discussed by Hillel or will come to mind when reading this stimulating book. □

N. W. Pirie is at the AFRC Institute of Arable Crops Research, Rothamsted Experimental Station, Harpenden, Hertfordshire AL5 2JP, UK.

Bad behaviour

Jeffrey Mervis

Impure Science: Fraud, Compromise and Political Influence in Scientific Research. By Robert Bell. Wiley: 1992. Pp. 301. \$22.95 (US and Canada only).

As examples of misconduct in science mount, it is natural for observers to try to weave common threads through them. But it is damnably difficult to find useful similarities among situations that range from the arid Rift Valley of Ethiopia to the most modern of university laboratories, and from the clanging assembly lines of a defence contractor to the hushed meetings of a scientific advisory panel.

Such is the problem that confronts Robert Bell, an economist at Brooklyn College, City University of New York. Bell wants desperately to say something important about his subject, and he is constantly imploring the reader to see the 'big picture' in his descriptions of misconduct. But he never quite manages himself to draw that picture.

Bell seems prepared to jettison the consensus definition of scientific fraud — falsification, fabrication and plagiarism of data — in favour of another terrible triumvirate more in tune with his politi-

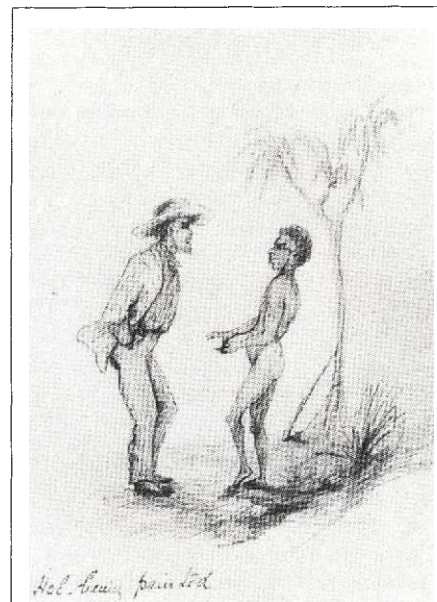
cal beliefs: 'big science', corporate support, and anything involving the US military establishment. It is an unhappy choice of targets, not because they are blameless, but rather because picking on them sheds so little light on the real causes of misconduct or on any potential solutions.

Bell explains that he has deliberately chosen cases from widely varying fields to show how pervasive is the stain of misconduct on the corpus of science. Yet one is left with the uncomfortable feeling that his selection process was arbitrary, reflecting more his personal prejudices and the availability of public source material than any real links between cases or concern for their larger relevance. How else to explain the lengthy treatment of the sad case involving Jon Kalb, a geologist working on prehistoric sites in Ethiopia who was unfairly black-listed by the National Science Foundation (NSF) after baseless rumours of his being an agent for the CIA were raised during a review of his grant proposal? Or the political fight that ensued after the NSF awarded a five-year, \$25 million grant for an engineering earthquake centre to the University of Buffalo rather than to the University of California, Berkeley?

It took Kalb, who wound up with \$20,000 and an apology from the NSF, more than 10 years to prove that he had been wronged, and his diligence uncovered a hidden filing system that the NSF has since abandoned. Bell quotes liberally from Kalb's files, but the best he can do by way of a conclusion is to say that "scientists must speak up when they feel that their proposals have been treated unfairly". Who would disagree?

In the case of the earthquake centre, Bell makes the hardly surprising discovery that politics plays a role in the awarding of large grants. And his conviction that the centre should have been built in California — "the best scientists ... were not funded, while people of lesser achievement were funded" — places him on very shaky ground indeed. He further claims, without substantiation, that the NSF pressured the National Academy of Sciences to restrain its criticism of a study of earthquake research that a US senator from California had convinced his congressional colleagues to fund as part of an unsuccessful campaign to reverse the NSF's original decision. This time his conclusion is not only weak, but also illogical: "If large science projects are not free of political influence, how can we expect smaller, less visible projects to be decided fairly and objectively?"

Such opinions could perhaps be excused if the author had marshalled them to prescribe a strong cure. Alas, he does not. His primary solution is to give



Participant observation — T. H. Huxley drew this sketch of himself having cosmetics applied by a "native acquaintance" probably while in Rockingham Bay in 1884 during his service on the HMS *Rattlesnake*. Huxley was central to the organization of anthropology as a discipline in Britain, and was committed to combatting racism and sexism. The picture is taken from *The Savage Within: The Social History of British Anthropology 1885–1945* by Henrika Kuklick. Published by Cambridge University Press, price £30, \$44.95.

whistle-blowers more incentive to speak out and more protection when they do, and to let the courts resolve scientific disputes. Even if one believes that the legal system is capable of meting out justice in cases of misconduct, it is hard to imagine how the verdicts could be used by the vast number of scientists who, in Bell's mind, need to be kept on the straight and narrow. He also suggests that those involved in obtaining federal monies for their institutions — from the president of a university to the principal research investigator — should be excluded from any inquiries into alleged misconduct on the grounds that they are compromised by a desire to perpetuate the present system. If that practice were followed, few scientists and administrators would be left to investigate misconduct, or to help other, 'purer' souls to carry out the necessary work.

An effective way to deal with allegations of misconduct must be found if science hopes to retain its favoured position as an enterprise worthy of public support. Unfortunately, appropriate remedies are not to be found here. □

Jeffrey Mervis is news editor of Nature.

Proteolysis, proteasomes and antigen presentation

Alfred L. Goldberg & Kenneth L. Rock

Proteins presented to the immune system must first be cleaved to small peptides by intracellular proteinases. Proteasomes are proteolytic complexes that degrade cytosolic and nuclear proteins. These particles have been implicated in ATP-ubiquitin-dependent proteolysis and in the processing of intracellular antigens for cytolytic immune responses.

PROTEINS in eukaryotic cells are continually degraded and replaced. Although this continual proteolysis may seem wasteful, it serves several important functions¹: cells selectively degrade proteins with abnormal conformations, the accumulation of which could be harmful; the rapid degradation of regulatory peptides is essential for the control of metabolic pathways and the cell cycle; and the breakdown of dispensable proteins in starvation provides amino acids for energy metabolism. Recent studies²⁻⁷ have established another critical function of cell proteases. In the development of the immune response, the presentation to T lymphocytes of protein antigens requires their intracellular digestion to small peptides, which are then bound to major histocompatibility complex (MHC) molecules and transported to the cell surface.

Proteolytic pathways

Protein breakdown in eukaryotic cells occurs through distinct pathways, which also have distinct roles in the processing of antigens for presentation to T lymphocytes. Extracellular antigens are taken up by endocytosis into MHC class II-expressing, antigen-presenting cells^{8,9}. These proteins are then degraded by the endosomal and lysosomal compartments by acid-optimal proteases (cathepsins)^{8,9}, which are also involved in the normal turnover of membrane proteins and in the accelerated degradation of cytosolic proteins in poor nutritional conditions¹⁰. These enzymes break down the endocytosed antigens to oligopeptides of about 13–18 residues, some of which bind to MHC class II molecules and are thereby transported to the plasma membrane. Eventually, the peptide-MHC complexes are recognized by helper T lymphocytes that stimulate inflammatory and antibody responses^{8,11,12}.

The present article focuses on the pathway by which peptide fragments are generated from antigenic proteins synthesized

inside the cell. Such peptides eventually bind to newly synthesized class I MHC molecules and are delivered to the cell surface for presentation to cytotoxic T lymphocytes. Those peptides to which the host is not tolerant, typically mutant or viral peptides, are recognized by the cytotoxic T lymphocytes which destroy the offending cell. Although MHC class I molecules bind peptides in the endoplasmic reticulum/vacuolar system^{13,14}, several observations suggest that these protein fragments are generated in the cytosol. First, their processing does not require an acidic compartment or lysosomal proteases¹⁵, and peptides produced in lysosomal compartments are generally not presented on class I molecules¹⁶. Second, antigens injected into the cytosol¹⁶, or that remain in the cytosol because they lack a signal sequence to target them into the lumen of the endoplasmic reticulum^{17,18}, are presented efficiently with class I molecules. Finally, the failure of some mutant cells to present antigen on class I molecules seems to be due to the absence of the transporter system that is believed to ferry peptides from the cytosol into the lumen of the endoplasmic reticulum where they bind to the newly synthesized class I molecules^{19,20}.

The proteolytic system that degrades the bulk of cell proteins, including the rapid elimination of highly abnormal proteins and short-lived regulatory polypeptides, is non-lysosomal and ATP-dependent (Fig. 1)^{1,8,21-23}. An initial step in the hydrolysis of most such proteins is a marking reaction involving their covalent conjugation to the polypeptide, ubiquitin²¹⁻²⁵. In this reaction, the carboxy terminus of ubiquitin becomes attached by an isopeptide bond to ϵ -amino groups on lysine residues on the protein substrate or on other ubiquitins. Much of the selectivity for proteolysis resides in this conjugation step, which is generally rate-limiting and requires ATP, at least three enzymes, and for certain polypeptides, a transfer RNA-dependent modification^{22,24}. Linkage to a chain of ubiquitins marks the polypeptide for rapid hydrolysis by a very large (26S or 1,300K) proteolytic complex, the ubiquitin-conjugate-degrading enzyme (UCDEN), whose function also requires ATP^{25-31,89} (see below). In addition, certain proteins are hydrolysed by an ATP-dependent process not involving ubiquitin³¹⁻³³, although the features distinguishing these proteins remain uncertain.

An essential component of these non-lysosomal pathways is the 20S proteasome particle (or multicatalytic protease complex)^{30,31,34-36}, and recent findings²⁻⁷ (see below) suggest a role for this structure in antigen presentation by MHC class I molecules. Thus, during evolution, the immune system seems to have used elements of this proteolytic pathway, which itself functions like an 'intracellular immune system' that screens and removes aberrant structures. Eukaryotic cells contain additional proteolytic systems in the endoplasmic reticulum³⁸ and in mitochondria³⁷ that cleave and degrade signal sequences and also

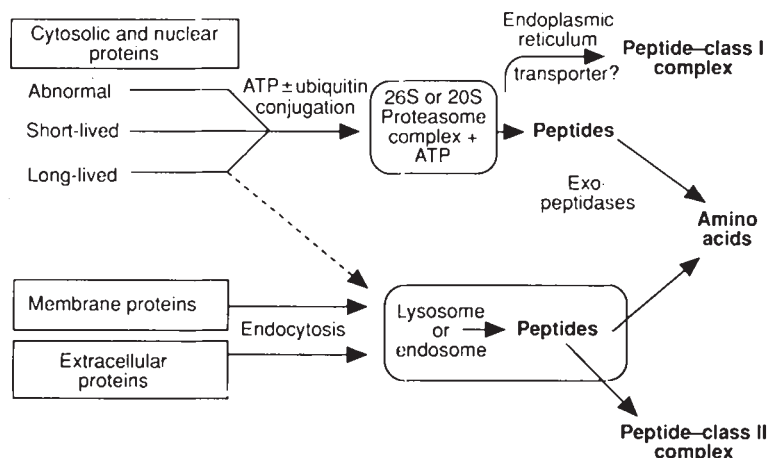


FIG. 1 Major pathways of protein degradation in eukaryotic cells and proposed roles in antigen presentation.

serve as a 'quality control system' for secreted or organellar proteins. Peptides generated in the endoplasmic reticulum from signal sequences are found on MHC class I molecules, even in mutant cells that seem not to transport peptides from the cytosol into the endoplasmic reticulum (see below). Thus, proteases in the endoplasmic reticulum can generate peptides for antigen presentation³⁹, although this process can only be a minor source of peptides presented. The ATP-dependent proteolytic system present in mitochondria³⁷ has not yet been implicated in antigen presentation, although proteins synthesized in this organelle can be presented on MHC class I molecules^{40,41}.

The 20S proteasome

The proteasome is a 700K particle with an essential role in the ATP-ubiquitin-dependent pathway⁴²⁻⁴⁴ and probably in other non-lysosomal degradative processes. This structure contains the major neutral proteolytic activity and is almost 1% of protein in mammalian cells (for reviews, see refs 31, 34-36). It is found in both the nucleus and cytosol of all eukaryotic cells⁴⁵⁻⁴⁸, and is composed of 13-15 distinct subunits of similar size (21-31K) but distinct isoelectric points³⁴⁻³⁶. These particles contain three or four different peptidase activities, which cleave specifically after basic, hydrophobic, or acidic residues^{34-36,47,48}. The subunits responsible for these activities are unknown, although mutants in a single subunit in yeast can inactivate a single catalytic activity⁵¹. Complementary DNAs encoding half of these subunits have now been cloned^{36,48-53}. These genes seem to be related to one another and are highly conserved from yeast to man; however, no homologies to known proteases or other gene families have been established³⁶.

It remains unclear why such a complicated structure containing about 30 subunits is necessary for intracellular proteolysis. These subunits are stacked into a barrel-shaped complex of four layers which on cross-section are ring-shaped (160 Å diameter) with a clear, central hole of 10-20 Å (refs 49, 54, 55), which may be a site of entry for protein subunits. Presumably, the large number of distinct subunits and their tight organization allow efficient protein breakdown and careful regulation of proteolytic activities *in vivo* (see below). Digestion in the barrel could be a valuable mechanism for allowing proteolysis to be processive and extensive, while preventing nonspecific proteolytic damage to the cell constituents. Very similar structures have been described in all eukaryotes examined^{34-36,49-51,55}. Moreover, archaeobacteria contain a homologous particle (but composed only of two subunits) that seem to be an evolutionary precursor of the eukaryotic proteasome (a 'proto-proteasome')⁵⁶. There may well be structural and functional differences amongst the proteasomes in a single cell, because these particles are found in nuclear^{45,46}, cytosolic, and microsomal fractions^{57,58}, as well as on the intermediate filaments of the cytoskeleton⁵⁹. Also, proteasomes vary in subunit composition during development^{49,60,61}, and recent studies suggest that a special type of proteasome is involved in antigen presentation (see below)²⁻⁷.

This particle has been studied in many different contexts, 21 different names have been assigned to it, and multiple functions proposed. A 600K ATP-activated protease was first reported in rat liver in 1979 and suggested to have a role in the ATP-dependent degradative pathway^{62,63}. A complex with multiple peptidase activities was first described in studies of pituitary peptide metabolism⁶⁴. These particles were also isolated as ribonucleoproteins called 'prosomes', which were believed to programme messenger RNA translation^{59,60}. But the reports of associated RNA species^{60,65} or of non-proteolytic functions remain controversial. 'Prosomes' and the multicatalytic protease were then shown to be identical by several criteria^{45,65,66}. The name 'proteasome' was therefore proposed to indicate its enzymatic and particular nature⁴⁵ and to simplify the literature. A more precise and preferable term would be the 20S proteasome

to distinguish this particle from the 26S proteasome complex, discussed below (Fig. 2).

The 26S proteasome complex

Although the 20S proteasome is essential in the ubiquitin-dependent pathway⁴²⁻⁴⁴, by itself, this particle cannot digest ubiquitin-conjugates and requires additional components for this process (see below). Studies by Rechsteiner^{25,28,29}, Goldberg^{26,27,31,89} and their colleagues demonstrated initially that ubiquitinated proteins are degraded to small peptides by a very large (26S or 1,300K) structure, which was called the 26S protease complex or UCEDN. But the purified structure not only degrades ubiquitin-conjugates by an ATP-dependent process but also some non-ubiquitinated proteins and short peptides. Therefore, we recently proposed the simpler name, 26S proteasome, to emphasize that it is a proteolytic particle distinct from but related to the 20S particle.

Recent studies have established that the proteasome is in fact an essential component of the larger structure⁴²⁻⁴⁴. Although early work had suggested that these particles had distinct functions^{26-29,89}, when the 20S protein was immunoprecipitated or chemically inactivated⁴², the ATP-ubiquitin-protein pathway and the breakdown of ubiquitin-conjugated proteins were blocked. A critical finding by Hershko and co-workers⁶⁷ was that when reticulocytes were depleted of ATP, no 26S complex could be found, but this structure could be formed *in vitro* by an ATP-dependent association of three smaller components, called CF-1 (600K), CF-2 (250K) and CF-3 (600K). Two groups^{43,44} then demonstrated that the 20S proteasome corresponds to component CF-3, and that in the presence of ATP, this particle associates with the other components to form the larger complex that degrades ubiquitinated proteins. Thus, both the formation of the 26S complex and its proteolytic function require ATP hydrolysis^{44,67,68} (see below). It is uncertain to what extent the 20S proteasomes exist and function *in vivo* as free particles or as components of the 26S complex. Similarly, it is unclear whether the formation of the larger complex is strictly a biosynthetic step or whether its formation and dissociation occur continually as part of a reaction cycle or whether these are regulatory events by which the cell controls overall rates of proteolysis.

The 26S complex contains the multiple peptidase activities of the 20S proteasome^{41,42} and probably all its subunits (21-31K) plus 10-20 larger polypeptides (40-110K)³⁶. Recently, clear images of the purified 26S complex have been obtained by negative staining^{69,70} and by a new freeze-fracture technique developed by J. Heuser⁹¹ (Fig. 3). Two forms are evident: One

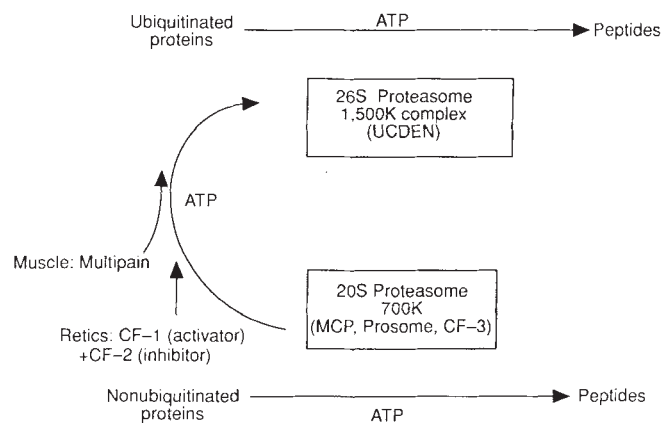


FIG. 2 Different forms of the proteasome and their functions in intracellular proteolysis. Formation *in vitro* of the 26S complex and ubiquitin-conjugate degradation requires ATP, the 20S proteasome, and either CF-1 plus CF-2 or multipain, the interrelationship between the latter factors is unknown, as are the conditions leading to dissociation of the 26S complex.

resembles a 'champagne cork' composed at one end of the 20S barrel-shaped proteasome but with a rhombohedral base, which presumably contains the components that allow recognition of ubiquitinated proteins (see below). These images partly resemble a model proposed by Rechsteiner²¹ based on structures described in cells 20 years ago by Shelton *et al.*⁷¹. Also present (Fig. 3) are larger forms in which these structures appear to contain a cap at the other end of the proteasome. The functional significance of these two forms is unclear.

Elucidation of the catalytic and regulatory functions of components, CF-1 and CF-2, in the 26S complex is likely to be of major importance (Fig. 3). Although the 20S proteasome seems to provide the proteolytic core of the 1,300K complex, CF-1 and CF-2 must provide the capacity for recognition of the ubiquitinated proteins as well as for the dissociation of poly-ubiquitin chains and release of free ubiquitin molecules when proteolysis is complete. Recently, we have succeeded in purifying CF-2 (ref. 72) and have shown that it is an ATP-stabilized inhibitor of the proteasome's peptidase activities^{72,73,90}. Thus, within the 26S complex, CF-2 seems to regulate proteolytic function apparently by an ATP-linked mechanism. In addition, we have purified a 600K complex that causes a dramatic activation of the proteasome's several peptidase activities (J. Frydman, J. Driscoll and A.L.G., manuscript in preparation). This activator also seems to bind ATP and to be a component of CF-1. Its actions are opposed by the 250K inhibitor, CF-2. These findings suggest that within the 26S complex, the 20S proteasome is normally maintained in an inactive state by the inhibitor, CF-2 (ref. 72). Presumably, the binding of ubiquitinated proteins or protein substrates releases this inhibition and allows the activator temporarily to stimulate proteolysis by an ATP-dependent mechanism. Such a model would be analogous to the specific activation of the bacterial ATP-dependent proteases by protein substrates^{30,31} (see below).

An additional functional component of the 26S proteasome was recently isolated from skeletal muscle (instead of ATP-depleted reticulocytes)^{74,75}. It is a 500K ATP-dependent protease, named multipain, and differs completely from the 20S proteasome in subunit composition, substrates degraded and inhibitor sensitivity. In the presence of ATP, multipain and the 20S proteasome (plus some associated proteins) form a 1:1 complex resembling the 26S particle⁷⁴. By itself, multipain (unlike the proteasome) degrades ubiquitin-conjugates slowly and releases free ubiquitin, but formation of the larger complex accelerates these processes and allows proteolysis to proceed to smaller products⁷⁵. Thus, in the 1,300K complex, multipain and the proteasome seem to function synergistically. It is unclear how multipain^{74,75} from muscle relates to the components (CF-1 and CF-2) isolated from reticulocytes (refs 67, 72 and J. Frydman, J. Driscoll and A.L.G., manuscript in preparation).

These different findings may be due to the ability of this large complex to disassemble in different ways depending on the mode of isolation. Alternatively, the proteolytic machinery in these cells may differ in structure. In muscle, this proteolytic pathway is regulated by diet, hormones and muscular activity⁷⁶, whereas in reticulocytes, this system catalyzes the destruction of most proteins during maturation into erythrocytes^{77,78,92}.

ATP and the mechanism of proteolysis

In size and complexity, the 26S proteasome approaches the ribosome, and a major challenge will be to elucidate in detail its multistep mechanism. One fundamental question is why ATP is necessary for proteolysis²⁶⁻³². ATP interacts with multiple components on this large complex (refs 67, 72, 74 and J. Frydman, J. Driscoll and A.L.G., in preparation). Moreover, this particle has significant ATPase activity⁶⁸ that can be activated by protein and peptide substrates (E. Folco, N. Tawa, H. Sawada, H. Sawada, M. Sawada and A.L.G., manuscript in preparation). These findings indicate a stoichiometric linkage between protein and ATP hydrolysis by the 26S complex. It has often been stated that the 20S proteasome does not require ATP. But when the 20S proteasome is gently isolated from muscle, it also needs ATP for proteolysis but this requirement is very labile on purification³². Thus, in its native state, the 20S proteasome probably also requires ATP for function.

Because traditional proteases do not show a requirement for ATP and because the hydrolysis of peptide bonds is thermodynamically favoured, the energy requirement for intracellular proteolysis has been extensively investigated^{1,21-24,30,31,79}. Studies of the ATP-requiring degradative system in reticulocytes led to the discovery by Hershko, Ciechanover and co-workers of the involvement of ubiquitin conjugation, which is an ATP-requiring process²³⁻²⁶ that provides specificity to the degradative pathway. Meanwhile, studies of this energy requirement in *Escherichia coli* and mitochondria, which lack ubiquitin, led to the discovery of a new type of protease, in which protein and ATP cleavage are linked^{30,31,79}. Knowledge about these large multimeric enzymes should provide insight into the ATP-dependence of the 26S proteasome. For example, the mechanism of protease La (*lon*), the best studied such enzyme, seems to have evolved to prevent excessive or nonspecific digestion of cell constituents^{30,31,80}. This bacterial enzyme is strongly inhibited under intracellular conditions, but the binding of an unfolded protein promotes ATP-binding and exposure of proteolytic sites. Subsequent cycles of ATP hydrolysis allows the enzyme to degrade processively the bound protein and then leave the protease in an inactive state^{30,31,80}.

A number of our recent observations suggest that similar mechanisms apply to the 26S proteasome; for example, its peptidase and ATPase are also activated by protein substrates

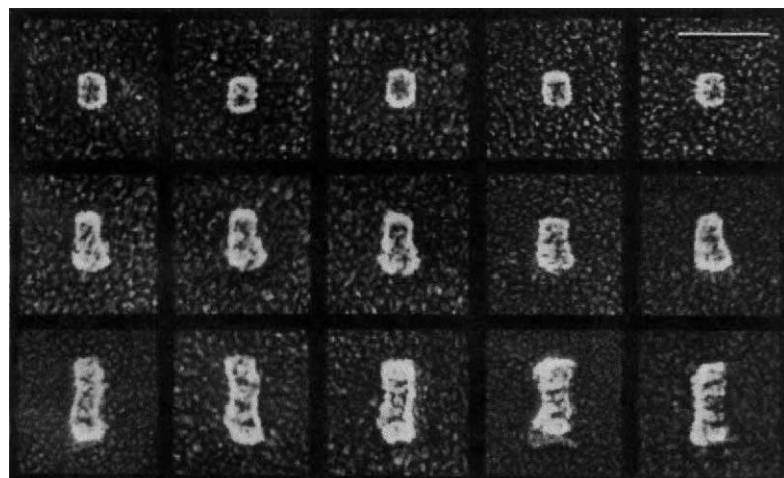


FIG. 3 Panel of freeze-dried and platinum-replicated proteolytic complexes from rabbit skeletal muscle (J. Heuser *et al.*, manuscript in preparation). The upper row displays the 20S proteasome and the centre row displays the 26S proteolytic complex (UCDEN), consisting of a proteasome on top connected to a rhombohedral expansion below. The third row displays larger complexes consisting of proteasomes capped at both ends by two rhombohedral expansions. The functional importance of these two forms is uncertain. Scale bar, 50 nm. These photomicrographs were kindly prepared by J. Heuser (Washington University, St Louis, Mo.) as described elsewhere⁹¹.

(E. Falco, N. Tawa, H. Sawada, M. Sawada and A.L.G., in preparation). The temporary activation of a latent 26S or 20S proteasome by substrates (such as ubiquitin conjugates) and linkage to ATP hydrolysis may help explain how these structures can exist in the nucleus or cytosol without causing widespread damage to cell proteins. Unlike the lysosome, proteasomes are not segregated in a membrane-enclosed space. Therefore, their activity must be tightly controlled. In fact, when the 20S particles are gently isolated, their various proteolytic activities are latent, although they can be activated dramatically by storage or exposure to various denaturants^{32,34,35,54} or the 600K activator. Such treatments may simply inactivate endogenous inhibitors and make the proteolytic sites more accessible to substrates. The prevention of such uncontrolled, potentially toxic activation of the proteasome (for example by the inhibitor CF-2 (refs 72, 73) or by making the structure specific to ubiquitinated proteins) would seem to be critical for cell viability.

Proteasome and antigen presentation

An exciting new development has been the recognition that the proteasome not only functions in the complete digestion of cytosolic proteins but may also be the proteolytic machinery that generates antigenic peptides for presentation on class I MHC. As pointed out by Parham⁸¹, the 20S proteasome resembles in size and subunit composition the MHC-linked low molecular mass proteins, LMP. This particle was discovered by Monaco and McDevitt when it was unexpectedly precipitated from cells with anti-MHC alloantisera⁸². These workers were first to speculate that it might have a role in antigen processing⁸³, but its precise function long remained obscure. Five recent reports²⁻⁷ have provided strong evidence suggesting that the LMP is a special type of 20S proteasome. (1) Anti-LMP and anti-proteasome antisera precipitate similar polypeptides². (2) Two subunits of the immunoprecipitated proteasomes show an allelic polymorphism on two-dimensional gels that maps in the class II region of the MHC locus^{2,5}. (3) Furthermore, the deduced protein sequence from two distinct cDNAs encoded in this MHC region are homologous to known proteasome subunits and presumably encode the polymorphic subunits^{3,4,6}. (4) Antisera against these two subunits precipitated only a small fraction of the proteasomes². (5) LMP particles have proteolytic activity (J. Monaco, personal communication).

These observations plus the location of these genes in the MHC region critical for antigen presentation indicate that these structures may be the cytosolic protease responsible for class I antigen processing. But direct evidence for such a role is lacking and neither LMPs nor proteasomes have been shown to generate the correct antigenic peptides (see below), either by themselves or as part of the 26S complex. It is not clear whether the MHC-encoded proteasome subunits are found in the 26S complex because the conditions used in these studies^{2,5} would have favoured dissociation of this structure. Townsend *et al.*⁸⁴ presented data suggesting a role of the ubiquitin-dependent degradative system in antigen presentation. As shown by Varshavsky and co-workers, proteins with large, charged amino termini are very rapidly degraded by this pathway^{24,85}. Genetic constructs where such termini are on a viral antigen enhance its presentation on MHC class I molecules⁸⁴. But evidence directly linking ubiquitin-conjugation or the 26S proteasome with antigen presentation is of obvious importance.

Also uncertain is the total number of proteasome subunits encoded in the MHC locus, but large chromosomal deletions in this region affect only two LMPs⁵. One important finding² is the apparent heterogeneity of the 20S proteasomes in a single cell, because most of these particles were still present in cell lysates after removal of the LMPs with anti-MHC antibodies. Thus, multiple species of proteasomes may exist in the same cell with distinct functions and regulation. In fact, treatment of cells with γ -interferon increases the expression of the MHC-encoded proteasome genes and content of LMP particles^{3,6,83}.

The ATP-ubiquitin dependent process is essential for viability, and the loss of some (but not all) proteasome subunit genes in yeast is lethal^{51,53}. By contrast, lymphoblastoid cell lines tolerate deletions of the MHC-linked proteasome genes from both chromosomes⁸⁶. Thus, these MHC-encoded components are likely to subserve a specific immunologic role rather than functioning in general protein breakdown.

It is interesting to consider what role the MHC-encoded components of the LMPs might perform if present in proteasomes specially adapted for supplying peptides to class I molecules. Recently, several antigenic oligopeptides bound to class I molecules have been isolated and sequenced^{87,88}. These studies indicate precise size constraints (of 8-10 amino acids) on the peptides that optimally bind to a specific class I molecule. It is important to determine whether all proteasomes or LPM subclasses can generate these same oligopeptides from proteins. The ATP-dependent proteolytic pathway degrade proteins rapidly to amino acids without an accumulation of polypeptide intermediates, apparently through the concerted action of the proteasome and cytosolic peptidases. Cells contain high amounts of such exopeptidases. Because peptides generated by the proteasome must escape these enzymes and be transported across the endoplasmic reticulum, mechanisms probably exist to rescue peptides from further degradation and/or to direct these molecules to the endoplasmic reticulum.

Recently, genes encoding the putative peptide transporters have been cloned; their sequences are homologous to known ATP-dependent membrane transport systems which contain the ATP-binding cassette (ABC) motif⁸¹. These genes are also encoded in the MHC, adjacent to the proteasome subunits, and mutations in them markedly impair the ability of cells to present endogenous antigens with class I molecules^{19,20}. These molecules are present in the membrane of the endoplasmic reticulum and *cis* golgi (J. Trowsdale, personal communication), although direct evidence that they transport peptides has not yet been obtained. The delivery of peptides to these putative transporters might be easily accomplished if a subset of proteasomes were physically associated with peptide transporters on the endoplasmic reticulum². In fact proteasomes have been found in microsomal fractions⁵⁷, and associated with the endoplasmic reticulum⁵⁸. It will therefore be necessary to determine whether the MHC-encoded components influence the subcellular localization of the proteasome or the nature of the peptides generated or their susceptibility to exopeptidases. Clearly, greater knowledge about the biochemical mechanisms and cellular roles of the 20S and the 26S proteasomes are important not only for cell biology, but also for a full understanding of antigen presentation. □

Alfred L. Goldberg is at the Department of Cellular and Molecular Physiology, Harvard Medical School; Kenneth L. Rock is at the Department of Pathology, Harvard Medical School and Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.

- Goldberg, A. L. & St John, A. C. *A. Rev. Biochem.* **45**, 747-803 (1976).
- Brown, M. G., Driscoll, J. & Monaco, J. *J. Nature* **353**, 355-357 (1991).
- Glynne, R. *et al. Nature* **353**, 357-360 (1991).
- Martinez, C. K. & Monaco, J. *J. Nature* **353**, 664-667 (1991).
- Ortiz-Navarrete, V. *et al. Nature* **353**, 662-664 (1991).
- Kelly, A. *et al. Nature* **353**, 667-668 (1991).
- Robertson, M. *Nature* **353**, 300-301 (1991).
- Allen, P. M., Babbitt, B. P. & Unanue, E. R. *Immunol. Rev.* **98**, 171-187 (1987).
- Puri, J. & Factorovich, Y. *J. Immun.* **141**, 3313-3317 (1988).
- Dice, J. F. *FASEB J.* **1**, 349-357 (1987).
- Rudensky, A. Y., Preston-Hurlburt, P., Hong, S.-C., Barlow, A. & Janeway, C. A. Jr *Nature* **353**, 622-627 (1991).
- Brodsky, F. M. *Nature* **348**, 581-582 (1990).
- Townsend, A. R. M. *et al. Nature* **340**, 443-448 (1989).
- Yewdell, J. W. & Bennink, J. R. *Science* **244**, 1072-1075 (1989).
- Morrison, L. A., Lukacher, A., Braciale, V. L., Fan, D. P. & Braciale, T. J. *J. exp. Med.* **163**, 903-921 (1986).
- Moore, M. W., Carbone, F. R. & Bevan, M. J. *Cell* **54**, 777-785 (1988).
- Townsend, A. R. M., Gotch, F. M. & Davey, J. *Cell* **42**, 457-467 (1985).
- Townsend, A. R. M., Bastin, J., Gould, K. & Brownlee, G. G. *Nature* **324**, 575-577 (1986).
- Spies, T. & Demars, R. *Nature* **351**, 323-324 (1991).
- Powis, S. J. *et al. Nature* **354**, 529-531 (1991).

21. Rechsteiner, M. A. *Rev. Cell Biol.* **3**, 1–30 (1987).
22. Hershko, A. *J. Biol. Chem.* **262**, 8303–8313 (1988).
23. Dice, J. F. *Seminars in Cell Biology* (ed. Dice, J. F.) (Saunders, London, 1990).
24. Finley, D. & Chau, V. A. *Rev. Cell Biol.* **7**, 25–69 (1991).
25. Hough, R., Pratt, G. & Rechsteiner, M. in *Ubiquitin* (ed. Rechsteiner, M.) 101–128 (Plenum, New York, 1988).
26. Waxman, L., Fagan, J. M. & Goldberg, A. L. *J. Biol. Chem.* **262**, 2451–2457 (1987).
27. Waxman, L., Fagan, J. M. & Goldberg, A. L. *Fedr. Proc.* **45**, 694 (1986).
28. Hough, R., Pratt, G. & Rechsteiner, M. *J. Biol. Chem.* **262**, 8303–8313 (1987).
29. Hough, R., Pratt, G. & Rechsteiner, M. *J. Biol. Chem.* **261**, 2391–2399 (1986).
30. Goldberg, A. L. in *Seminars in Cell Biology* (ed. Dice, J. F.) 423–432 (Saunders, London, 1990).
31. Goldberg, A. L. *Eur. J. Biochem.* **203**, 9–23 (1992).
32. Driscoll, J. & Goldberg, A. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 787–791 (1989).
33. Bercovitch, Z., Rosenberg-Hassan, Y., Ciechanover, A. & Kahana, C. *J. Biol. Chem.* **264**, 15949–15952 (1989).
34. Orlowski, M. *Biochemistry* **29**, 10289–10297 (1990).
35. Rivett, A. *J. Archs Biochem. Biophys.* **218**, 1–8 (1989).
36. Tanaka, K., Tamura, T., Yoshimura, T. & Ichihara, A. *New Biol.* **4**, 1–11 (1992).
37. Desautels, M. & Goldberg, A. L. *J. Biol. Chem.* **257**, 11673–11679 (1982).
38. Klausner, R. D. & Sittia, R. *Cell* **62**, 611–614 (1990).
39. Henderson, R. A. *et al. Science* **255**, 1264–1266 (1992).
40. Loveland, B. *et al. Cell* **60**, 971–980 (1990).
41. Davies, J. D. *et al. J. exp. Med.* **173**, 823–832 (1991).
42. Matthews, W., Driscoll, J., Tanaka, K., Ichihara, A. & Goldberg, A. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2586–2601 (1989).
43. Eytan, E., Ganoth, D., Armon, A. & Hershko, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7751–7755 (1989).
44. Driscoll, J. & Goldberg, A. L. *J. Biol. Chem.* **265**, 4789–4792 (1990).
45. Arrigo, A.-P., Tanaka, K., Goldberg, A. L. & Welch, W. J. *Nature* **331**, 192–194 (1988).
46. Tanaka, K., Kumatori, A., Kunio, I. & Ichihara, A. *J. Cell. Physiol.* **139**, 34–41 (1989).
47. Orlowski, M., Cardozo, C., Hidalgo, M. C. & Michaud, C. *Biochemistry* **30**, 599–6005 (1991).
48. Mykles, D. L. & Haire, M. F. *Archs Biochem. Biophys.* **288**, 543–551 (1991).
49. Scherrer, K. *Molec. Biol. Rep.* **14**, 1–9 (1990).
50. Haass, C., Pesold-Hurt, B., Muelhaup, G., Beryreuther, K. & Kloetzel, P. M. *EMBO J.* **8**, 2373–2379 (1989).
51. Heinemeyer, V., Kleinschmidt, J. A., Saidowsky, J., Escher, C. & Wolf, D. H. *EMBO J.* **10**, 555–562 (1991).
52. Tanaka, K. *et al. Biochemistry* **29**, 3777–3785 (1990).
53. Fujiwara, T. *et al. J. Biol. Chem.* **265**, 16604–16613 (1990).
54. Tanaka, K. *et al. J. molec. Biol.* **203**, 985–996 (1988).
55. Kopp, F. *et al. Biochim. Biophys. Acta* **872**, 253–260 (1986).
56. Zwickl, P., Lottspeich, F., Dahlmann, B. & Baumeister, W. *FEBS Lett.* **278**, 217–221 (1991).
57. Tanaka, K., Li, K., Ichihara, A., Waxman, L. & Goldberg, A. L. *J. Biol. Chem.* **261**, 15197–15203 (1986).
58. Knecht, E., Palmer, A., Sweeney, S. T. & Rivett, A. *J. Biochem. Soc. Trans.* **19**, 2935 (1991).
59. Grassi de Sa, M.-F. *et al. J. Cell. Biol.* **107**, 1527 (1988).
60. Gounon, M. F., Grassi de Sa, M.-F. & Scherrer, K. *J. Cell Sci.* **90**, 555 (1988).
61. Ahn, J. Y. *et al. J. Biol. Chem.* **266**, 15746–15749 (1991).
62. DeMartino, G. N. & Goldberg, A. L. *J. Biol. Chem.* **254**, 3712–3715 (1979).
63. Rose, I. A., Warmis, J. D. B. & Hershko, A. *J. Biol. Chem.* **254**, 8135–8138 (1979).
64. Wilk, S. & Orlowski, M. *J. Neurochem.* **35**, 1172–1182 (1980).
65. Falkenburg, P. E. *et al. Nature* **331**, 90–92 (1988).
66. Kleinschmidt, J. A., Escher, C. & Wolf, D. H. *FEBS Lett.* **239**, 35–42 (1988).
67. Ganoth, D., Leshinsky, E., Eytan, E. & Hershko, A. *J. Biol. Chem.* **263**, 12412–12419 (1988).
68. Armon, T., Ganoth, D. & Hershko, A. *J. Biol. Chem.* **265**, 20723–20726 (1990).
69. Peters, J.-M., Harris, J. R. & Kleinschmidt, J. A. *Eur. J. Cell Biol.* **56**, 422–432 (1991).
70. Ikai, A., Nishigai, M., Tanaka, K. & Ichihara, A. *FEBS Lett.* **292**, 21–24 (1991).
71. Shelton, E., Kuff, E., Maxwell, E. S. & Harrington, J. T. *J. Cell Biol.* **48**, 1–8 (1970).
72. Driscoll, J., Frydman, J. & Goldberg, A. L. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4986–4990 (1992).
73. Murakami, K. & Etlinger, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7588–7592 (1986).
74. Estrela, J. & Goldberg, A. L. *J. Biol. Chem.* (in the press).
75. Estrela, J. & Goldberg, A. L. *J. Biol. Chem.* (in the press).
76. Kettelhut, I. C., Wing, S. S. & Goldberg, A. L. *Diab. Metab. Rev.* **4**, 751–772 (1988).
77. Boches, F. & Goldberg, A. L. *Science* **215**, 1027–1033 (1982).
78. Speiser, S. & Etlinger, J. *J. Biol. Chem.* **257**, 14122–14127 (1982).
79. Gottesman, S. A. *Rev. Genet.* **23**, 163–198 (1989).
80. Menon, A. S. & Goldberg, A. L. *J. Biol. Chem.* **262**, 14929–14934 (1987).
81. Parham, P. *Nature* **348**, 674–675 (1990).
82. Monaco, J. J. & McDevitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3001–3005 (1982).
83. Monaco, J. J. & McDevitt, H. O. *Hum. Immun.* **15**, 416–426 (1986).
84. Townsend, A. *et al. J. exp. Med.* **168**, 1211–1224 (1988).
85. Bachmair, A., Finley, D. & Varshavsky, A. *Science* **234**, 179–186 (1986).
86. Spies, T. *et al. Nature* **348**, 744–747 (1990).
87. Van Bleek, G. M. & Nathenson, S. G. *Nature* **348**, 213–216 (1990).
88. Falk, K., Rotzschke, O. & Rammensee, H.-G. *Nature* **348**, 248 (1990).
89. Fagan, J. M., Waxman, L. & Goldberg, A. L. *Biochem. J.* **243**, 335–343 (1987).
90. Li, X., Gu, M. & Etlinger, J. D. *Biochemistry* **30**, 9709–9715 (1991).
91. Heuser, J. E. *J. Elect. Micro Meth.* **13**, 244–263 (1989).
92. Muller, M., Dubiel, W., Rothman, J. & Rapoport, S. *Eur. J. Biochem.* **109**, 405–410 (1980).

ACKNOWLEDGEMENTS. We are grateful to Mrs Aurora Scott for expert assistance in the preparation of this manuscript. Work from our laboratories discussed in this review was kindly supported by grants from the NIH, NASA, MDA and Harriett Ames Charitable Trust.

Is beryllium in metal-poor stars of galactic or cosmological origin?

Gerard Gilmore*, Bengt Gustafsson†, Bengt Edvardsson† & Poul E. Nissen‡

* Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

† Astronomical Observatory, Box 515, S-751 20 Uppsala, Sweden

‡ Institute of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark

Standard Big Bang nucleosynthesis predicts a very small primordial abundance of beryllium. Observations of nine very metal-poor stars indicate a beryllium abundance roughly proportional to the oxygen abundance, a trend that can be explained in terms of galactic chemical evolution. Combining this rate of beryllium production with recent observations of boron and lithium in similar stars yields an upper limit to the primordial beryllium abundance several orders of magnitude greater than the cosmological prediction, a result that can be explained by cosmic-ray activity in the early Galaxy.

THE abundances of the light elements created by primordial nucleosynthesis in the early Universe are a test of the essential assumptions that underlie the standard model of the Big Bang, and can in principle be used to deduce the total baryonic mass of the Universe. The most common light elements, ^1H , ^2H , ^3He , ^4He and ^7Li , have relative primordial abundances in adequate agreement with the standard model¹. The baryonic mass of the Universe derived from these abundances is in the range $0.04 \leq \Omega_B h_{50}^2 \leq 0.08$, where $\Omega_B = 1$ is the critical mass density such that the mass density in the corresponding model universe just balances the expansion, with the net energy content of that model being zero, and h_{50} is the Hubble constant in units of $50 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

The results are based on the simplifying assumption that the Universe was homogeneous when the initial conditions for nucleosynthesis were established, ~ 1 s after the Planck time.

But recent studies of the quantum chromodynamics (QCD) phase transition from quark–gluon domination to baryon domination of the Universe, which occurred $\sim 20 \mu\text{s}$ after the Planck time, suggest that the Universe may have been very inhomogeneous at early times. If the QCD phase transition is first order, hadronic matter comes into existence by growing in lumps around nucleation centres. As these lumps cool below $\sim 1 \text{ MeV}$, at a time of ~ 1 s, the weak interaction will cease to be in equilibrium, and neutrons will diffuse out of the high-density regions. Protons will be partially trapped by the $e^\pm$ plasma, causing variations in both local density and the local neutron/proton ratio², leading perhaps to observable effects on light element production, and to a potentially very large effect on the deduced value of Ω_B . There are arguments that both support³ and oppose⁴ identification of the phase transition as one of first order, and a variety of other possible causes for

TABLE 1 Model parameters and abundance results

Star	T_{eff} (K)	$\log g$	ξ_t	[Fe/H]	Date*	[O/H]†	$\log (N_{\text{Be}}/N_{\text{H}})$	$\log (N_{\text{Li}}/N_{\text{H}})$
HD16031	5,970	3.9	1.5	-1.96	9/91	-1.32	-12.37	-9.97
HD34328	5,730	4.6	1.1	-1.87	9/91	-1.46	-11.9	-9.93
HD84937	6,090	4.0	1.5	-2.41	4/91	-1.87	≤ -12.6	-9.95
HD116064	5,770	3.8	1.7	-2.18	4/91	-1.74	-12.70	
HD140283	5,540	3.5	1.5	-2.77	6/90	-2.10	-12.85	-10.02
HD140283					4/91	-2.10	-13.03	
HD160617	5,800	3.8	1.7	-2.06	4/91	-1.69	-12.38	-9.80
HD160617					9/91	-1.66	-12.55	
HD200654	5,090	2.7	1.5	-3.0‡	4/91	-2.26	≤ -13.7	
HD200654					9/91	-2.31	≤ -13.7	
HD213657	5,860	3.8	1.5	-2.28	9/91	-1.65	-12.65	-9.83
CD - 30° 18140	6,130	4.0	1.7	-2.11	9/91	-1.35	-12.28	-9.74
HR3018§	5,820	4.4	1.2	-0.78	9/91	-0.44	-11.20	
HR8181§	6,140	4.3	1.6	-0.67	9/91	-0.53	-11.40	

Lithium abundances are from ref. 28, and F. Spite and M. Spite (personal communication).

* Observing date for the ultraviolet spectrum (month/year).

† Oxygen abundances derived by P.E.N., B.E., B.G. and G.G. (manuscript in preparation).

‡ [Fe/H] estimated from Strömgren photometry.

§ Model parameters and iron abundance results from B.E., J. Andersen, B.G., D. L. Lambert, P.E.N. and J. Tomkin (manuscript in preparation).

inhomogeneity in the very early Universe have been discussed. Observational cosmology might provide further information concerning this question.

The physical details of the quark-hadron phase transition(s) remain uncertain, and calculations of the predicted abundances of the (potentially) measurable light elements (^2H , ^3He , ^4He , ^6Li , ^7Li , ^9Be , ^{10}B and ^{11}B) in inhomogeneous models disagree. Recent work⁵ indicates that back-diffusion of neutrons into the proton-rich regions should considerably reduce the nucleosynthetic effects of the inhomogeneities.

The ^9Be abundance expected in homogeneous standard Big Bang models is also uncertain; the most recent calculation for a Universe with $\Omega_{\text{B}} \approx 0.04$, the value consistent with available data¹, is $(N_{\text{Be}}/N_{\text{H}}) \approx 10^{-18}$ (ref. 6). The primordial boron to beryllium ratio in this calculation is $(N_{\text{B}}/N_{\text{Be}}) \geq 10$.

Thus, there is no guarantee that primordial ^9Be is detectable even in a fairly inhomogeneous Universe. If, however, a high primordial ^9Be abundance were detected, the homogeneous standard Big Bang model would require substantial modification, parameter space for all models of primordial nucleosynthesis would be severely restricted, the baryon density Ω_{B} would be known more reliably, and the physics of the Universe before the end of the quark-dominated era would be available for study. We have recently measured the abundance of Be in a very metal-poor star⁷, showing that such measurements are indeed possible. We have now extended our measurements to include eight very metal-poor stars in which ^7Li , another primordial light element whose evolution in the Galaxy is similar to that of ^9Be , is presumably at its primordial value. We use these data here to investigate whether there are any indications of a primordial abundance of ^9Be in excess of the value predicted in the homogeneous standard model.

Observations and data analysis

At temperatures of a few million kelvin, at the bottom of deep convective zones in cool stars, ^7Li and ^9Be are destroyed, ^7Li being the more fragile⁸. To be sure that the present atmospheric ^9Be abundance in a star is that of the interstellar medium from which it formed, we must ensure that the star has never had a deep convective zone in its upper atmosphere. Beryllium-9 is produced by spallation of heavy nuclei, in particular oxygen, by cosmic rays, so to understand the galactic production of ^9Be we need to know the oxygen abundance of the interstellar gas in the early Galaxy. Fortunately, that abundance can be derived from the spectra that are used to derive the ^9Be abundance, making use of OH molecular lines. We must also choose stars in which the ^9Be created by galactic processes after primordial nucleosynthesis is minimized, so we restrict attention to very

metal-poor stars whose ^7Li abundance is at the primordial value⁹ of $(N_{\text{Li}}/N_{\text{H}}) \approx 1.6 \times 10^{-10}$. This restricts us to unevolved stars with $T_{\text{eff}} > 5,500$ K, and with $[\text{Fe}/\text{H}] \leq -1.5$. ($[\text{Fe}/\text{H}] \equiv \log(N_{\text{Fe}}/N_{\text{H}})_{\text{star}} - \log(N_{\text{Fe}}/N_{\text{H}})_{\text{Sun}}$).

The only ^9Be lines accessible for analysis are the Be II lines at 3,131 Å. We have observed the spectral region from 3,100 to 3,200 Å for eight unevolved stars with $-2.8 \leq [\text{Fe}/\text{H}] \leq -1.9$, one somewhat evolved (subgiant) star with $[\text{Fe}/\text{H}] = -3.0$, and two unevolved stars with $[\text{Fe}/\text{H}] \approx -0.75$, in two observing runs (April 1991 and September 1991) with the 3.9-m Anglo-Australian Telescope. Observations at visual wavelengths to determine abundances of Mg, Ca, Ti, Cr and Fe in the same stars were made at the European Southern Observatory 3.6-m telescope. Complementary high-precision photometric data are also available¹⁰, allowing derivation of the effective temperature, T_{eff} , the surface gravity, g , the microturbulence parameter, ξ_t , and the overall metal abundance, $[\text{M}/\text{H}]$, for each star. Some details of the stars observed are given in Table 1.

The spectra were analysed using state-of-the-art model atmosphere techniques⁷ which will be described in detail elsewhere (P.E.N., B.E., B.G. and G.G., manuscript in preparation; B.E., J. Andersen, B.G., D. L. Lambert, P.E.N. & J. Tomkin, manuscript in preparation). We used synthetic spectra in the ultraviolet and calculated individual equivalent widths in the visual spectral region. Figure 1 shows examples of observed ultraviolet spectra and fits of synthetic spectra with different assumptions of beryllium abundances. Table 1 shows the resulting $[\text{Fe}/\text{H}]$ values of our stars (on a scale with solar $\log(N_{\text{Fe}}/N_{\text{H}}) = -4.50$; refs 11, 12) as derived from neutral iron lines (P.E.N., B.E., B.G. and G.G., manuscript in preparation). It also gives values for the main stellar atmospheric parameters.

Analysis of the abundances determined from different ionization states of iron suggests that the differential logarithmic values of stellar surface gravity have been determined to an accuracy better than ± 0.20 . Similarly, the absolute iron abundance values may be underestimated by ~ 0.15 dex (1 dex is one unit on the logarithmic scale), but the differential values have errors less than 0.08 dex: the differential values of surface gravity and iron abundance are most important for the present analysis.

Table 1 gives the abundances. The beryllium abundance of HD34328 is uncertain because of an uncertain spectrograph setting caused by problems with the observational equipment. The low upper limit abundance of HD200654 is not surprising, because the star is so cool that we expect it to be depleted in Li and Be (ref. 13).

The uncertainties in the Be abundances mainly come from the uncertainties in the surface gravities of the stars and in the continuum levels of the observed spectra. The relative values of

the logarithmic surface gravities of the stars are probably correct within ± 0.2 , causing an uncertainty of a factor of ~ 1.25 in the derived abundance. The continuum level in the first spectrum of HD140283, observed in June 1990, is more uncertain than those of the other spectra, taken in 1991. This is the main cause of the difference between the Be abundance derived from this spectrum and that derived from the March 1991 spectrum of the same star⁷; these are respectively, $\log(N_{\text{Be}}/N_{\text{H}}) = -12.85$ and -13.03 . The difference is larger than the continuum uncertainties appropriate for the later observations, so we give the first result half the weight of the latter in our mean value, -12.97 , for the star. Ryan *et al.*¹⁶ derived for this same star $\log(N_{\text{Be}}/N_{\text{H}}) = -13.2$, based on spectra observed with exactly the same equipment. The abundance difference of 0.23 in the log is mostly due to the lower gravity, $\log g = 3.2$, adopted by Ryan *et al.* Allowing for that systematic difference, the two measurements are in excellent agreement.

Typical errors in the Be abundances, including continuum and fitting uncertainties as well as uncertainties in the synthetic spectrum calculations, are estimated to be ± 0.25 in the logarithmic abundances. The scatter in the ratio of $[\text{Be}/\text{H}]$ to $[\text{Fe}/\text{H}]$, discussed in the next section, seems to indicate that this is a conservative estimate of the true abundance errors. The only serious systematic cause of error that we can imagine would be unidentified lines blending with Be in the stellar spectra; we consider this highly improbable but cannot completely rule it out. The resulting Be abundances and upper limits are given in Table 1.

Galactic evolution of beryllium

Beryllium-9 is created in primordial nucleosynthesis, and also in the destruction (spallation) of C, N and O nuclei by protons with energy ≥ 10 MeV. Such protons are found in cosmic rays, which are thought to be accelerated by supernovae. Carbon and oxygen, and perhaps nitrogen, are created in the supernova explosions of stars more than about 10 times as massive as the Sun. The supernovae that created elements heavier than ${}^4\text{He}$ in the very metal-poor stars considered here will have accelerated the cosmic-ray protons to high energies and provided the target CNO atoms, so that ${}^9\text{Be}$ could be produced by spallation. Similar arguments apply to B and to Li, so that at least some fraction of the observed abundance of the light elements must be due to galactic production. The abundance ratios due to cosmic-ray spallation differ from the primordial production ratios, so that one may hope to disentangle the two contributions. Of ${}^7\text{Li}$, ${}^9\text{Be}$ and ${}^{11}\text{B}$, ${}^7\text{Li}$ is the most easily destroyed in stellar interiors, so its abundance can be used to limit any destruction of ${}^9\text{Be}$. The abundance distribution of ${}^7\text{Li}$ in metal-poor stars, including some of those studied here, indicates a primordial ${}^7\text{Li}$ abundance in accord with the value predicted by the homogeneous standard Big Bang model^{18,9}, $(N_{\text{Li}}/N_{\text{H}}) \sim 1.5 \times 10^{-10}$. The mean measured ${}^7\text{Li}$ abundance for the stars in Table 1 is 1.2×10^{-10} , with a dispersion of ~ 0.12 dex (32%), so the stars of interest here are either at the primordial abundance or show some slight depletion. Their ${}^9\text{Be}$ abundances may then also be treated as retaining the abundance of the gas from which the star formed.

Some models of Li evolution in the Galaxy allow a high primordial production, $(N_{\text{Li}}/N_{\text{H}}) \sim 7 \times 10^{-9}$, with later stellar destruction. This high destruction rate would also reduce the original Be abundance, which must have been a factor of a few higher than at present.

The evolution in the early Galaxy of the light elements created by spallation^{8,14-23} is described by

$$\frac{d(N_{\text{x}}/N_{\text{H}})}{dt} \propto \zeta(t) \times \int \sigma(E) \times \Phi(E, t) dE - \text{loss terms} \quad (1)$$

where N_{x} is the abundance of the element being created through spallation of CNO nuclei, N_{H} the hydrogen abundance, $\zeta(t)$ the relative abundance of the target heavy elements, $\sigma(E)$ the reaction cross-section, which depends on the energy E of the cosmic rays, and $\Phi(E, t)$ the time-dependent cosmic-ray flux. $\sigma(E)$ is measured in a laboratory and for present purposes can be assumed known. The most important target element for the production of Be and B is oxygen, but stellar abundances are most conveniently defined in terms of the iron abundance $[\text{Fe}/\text{H}]$. For the stars of Table 1 oxygen abundances were measured directly and, except for the two most metal-rich stars, are consistent with a simple constant relation $[\text{O}/\text{Fe}] = 0.56$. Oxygen abundances are not yet available for the other metal-rich disk stars discussed below, but they can be estimated from the mean relationship between $[\text{Fe}/\text{H}]$ and $[\text{O}/\text{Fe}]$ over the relevant abundance range (see Table 1). For the cosmic-ray production of ${}^7\text{Li}$, the low-energy production is dominated by ${}^4\text{He} + {}^4\text{He}$ reactions^{17,18}. Because ${}^4\text{He}$ is an abundant primordial element, $\zeta(t)$ for Li production is effectively constant, and the ${}^7\text{Li}$ production rate depends only on $\Phi(E, t)$.

The target heavy elements, particularly oxygen, are created in massive supernovae, so $\zeta(t)$ is a measure of the time-integrated number of supernovae in the relevant part of the proto-Galaxy. Acceleration of cosmic rays is also probably due to winds and supernovae from high-mass stars, and the cosmic-ray flux before any losses may therefore also be parametrized by the heavy element abundance, $\zeta(t)$. For present purposes, the relevant volume must be a region in which target nuclei and cosmic rays are found (that is, supernovae have occurred) and in which the enriched gas is available for continuing star formation, so that the light elements are preserved for later study.

There are two possible limiting cases of spallation production. In the first case spallation takes place predominately in the immediate neighbourhood of supernovae in regions which are sufficiently enriched by heavy elements that the production of light elements by spallation is independent of the mean heavy-element abundance in the Galaxy, but is limited by the short-lived cosmic-ray energy density in the immediate vicinity. The resulting abundance of beryllium and boron should then be roughly proportional to the abundance of oxygen.

In the second limiting case the spallation takes place mainly in the interstellar medium (ISM) beyond the immediate vicinities of supernovae, and is limited by the availability of both target nuclei and long-lived cosmic rays. One would then expect the abundance of spallation products to be proportional to the

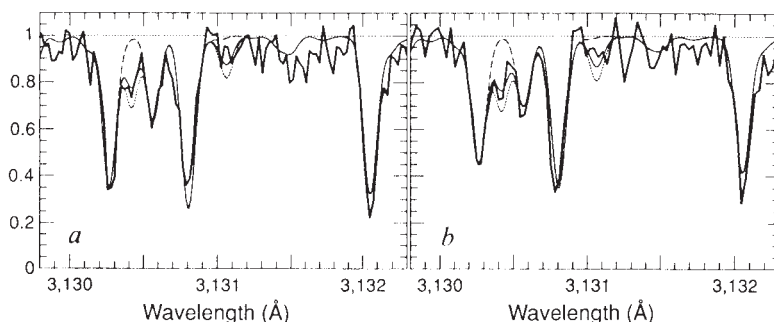


FIG. 1 Observed and synthetic spectra for two programme stars. *a*, HD116064; *b*, HD213657. The observed stellar spectra are shown by thick solid lines. The three thin lines show synthetic spectra with different assumed beryllium abundances: synthetic spectra with no beryllium absorption are shown as dashed lines, the thin solid lines shows synthetic spectra with the best-fit beryllium abundances (see Table 1), and the dotted lines show spectra with 0.2 dex (60%) higher beryllium abundances.

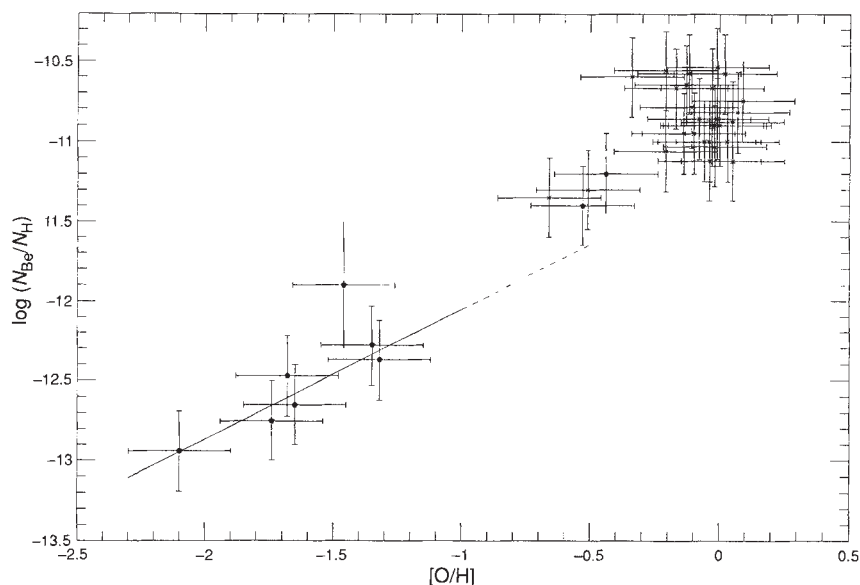


FIG. 2 Beryllium abundance against logarithmic oxygen abundance relative to the solar value, $[O/H]$ for programme stars (solid symbols), and other stars of Table 2 (crosses). The line is a least-squares fit to the high-quality data for stars with $[O/H]$ less than -1 (HD34328 excluded), and has slope 0.81, so that $N_{\text{Be}} \propto N_{\text{O}}^{0.8}$. This roughly linear proportionality is explained in the text.

square of the target abundance. The quadratic behaviour mainly comes from the two factors $\zeta(t)$ and $\Phi(E, t)$ in equation (1), both depending on the total number of supernovae up to time t . This is the case usually discussed. In galactic evolutionary models including these limiting cases in a more general framework (S. Feltzing and B.G., manuscript in preparation), estimates of relevant timescales indicate that the first case, with a galactic Be production proportional to the oxygen production, is realistic for the early Galaxy.

These two limiting cases are highly idealized but indicate how the abundance of a spallation product may depend on the abundance of its parent nuclei (created in supernovae, specifically of oxygen) to some power between 1 and 2.

Figure 2 shows the relationship between $[Be/H]$ and $[O/H]$ for our data of Table 1 and for all other published high-resolution data. The latter data are collected in Table 2, where we have also included new $[Fe/H]$ values and estimated $[O/H]$ values for all the stars. We have fitted weighted least-squares straight lines to the data shown in Fig. 2. Restricting the fit to the six stars with high-quality Be abundances, and with $[O/H] \leq -1$, gives a slope of 0.8. For all stars with $[O/H] \leq -0.4$ the

least-squares slope is 1.1, with a marginally better fit from a quadratic. The difference between these two slopes illustrates that by $[O/H] \approx -0.5$ the beryllium production is rising steeply to reach its disk level. For the metal-poor stars, however, the Be and O abundances are roughly proportional. We have also calculated a least-squares fit to the abundance data shown in Fig. 2 between $N_{\text{Be}}/N_{\text{H}}$ and $N_{\text{O}}/N_{\text{H}}$: a linear fit is acceptable only for stars with $[O/H] \leq -1$, and has slope 1.05. The Be abundance at zero O abundance is in this case 4×10^{-14} , but the standard Big Bang model abundance of $N_{\text{Be}}/N_{\text{H}} \approx 10^{-18}$ is allowed within the uncertainties, and provides a slope of 1.3.

A cosmological Be abundance?

Convincing evidence for a non-standard primordial abundance of ${}^9\text{Be}$ would be the detection of a plateau in the relationship between $[Be/H]$ and $[O/H]$. No such plateau has yet been detected, even though the most metal-poor star we have observed, HD140283, is more deficient in heavy elements than $\sim 95\%$ of the halo stars in the Milky Way. We now wish to quantify the relative production of light elements by spallation, to constrain the energy spectrum of cosmic rays in the proto-

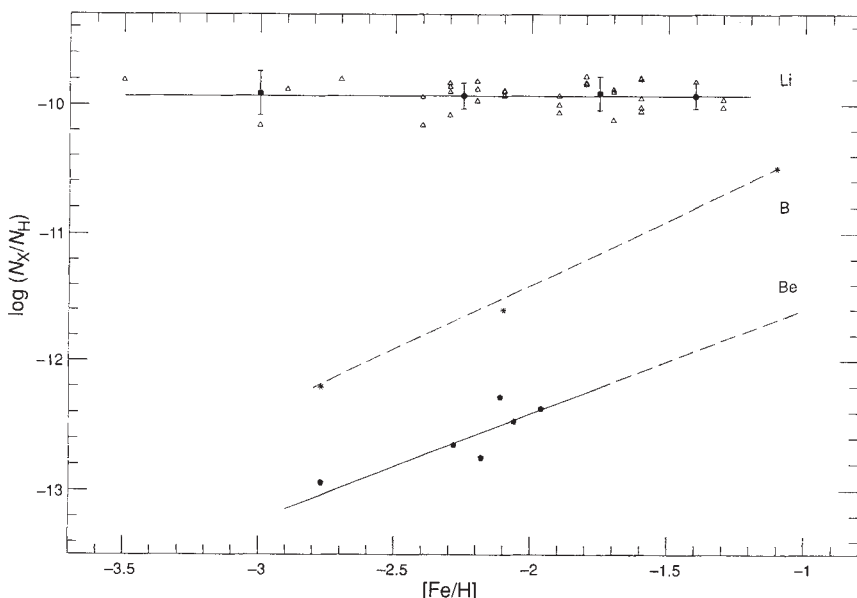


FIG. 3 Observed abundances of Li, Be and B in metal-poor stars (X represents Li, Be, B). The triangles denote lithium abundances measured for individual stars, and the solid dots represent lithium mean values for all stars in the metallicity intervals $[Fe/H] \leq -2.5$, $-2.5 < [Fe/H] \leq -2.0$, $-2.0 < [Fe/H] \leq -1.5$ and $-1.5 < [Fe/H] \leq -1.3$, with errors in these means indicated. The horizontal solid line indicates a halo lithium abundance of -9.93 . The line drawn through the beryllium data is the least-squares fit shown in Fig. 2, transformed from oxygen to iron abundance using the mean observed oxygen to iron ratio of the six stars studied here. The boron abundance estimates (asterisks) were obtained by D. Duncan, D. L. Lambert and M. Lemke (manuscript in preparation).

Galaxy and to test the extrapolation back to look for a possible primordial abundance. To do this, we use published observations of $[\text{Li}/\text{H}]$ against $[\text{Fe}/\text{H}]$, and measurements of the B/Be ratio at low abundances, to constrain the cosmic-ray energy spectrum $\Phi(E)$. We wish to determine whether there is a cosmic-ray energy spectrum that is consistent with all the observations, and whether enhanced primordial production of ${}^9\text{Be}$ and of ${}^{10}\text{B} + {}^{11}\text{B}$ is required.

The relative spallation production of Li, Be and B depends on a convolution of the energy-dependent cross-sections with the energy-dependent cosmic-ray spectrum, together with the abundance of target nuclei^{6,16,18,24,25}. The cross-sections depend on energy in such a way that the ${}^9\text{Be}$ production has a minimum near ~ 200 MeV, near the peak of the present-day cosmic-ray energy distribution. For cosmic-ray energies above ~ 300 MeV, the relative production rates are independent of particle energy. The production ratio $N_{{}^{10}\text{B}+{}^{11}\text{B}}/N_{{}^9\text{Be}}$ increases rapidly at very low energies (≤ 30 MeV) as resonances dominate the cross-sections. This energy dependence means that we can provide ratios over the range $10 \leq N_{{}^{10}\text{B}+{}^{11}\text{B}}/N_{{}^9\text{Be}} \leq 30$ by adopting suitably narrow-band or very steep power-law cosmic-ray energy spectra (ref. 14, and D. Lambert, personal communication). Integration of the energy-dependent cross-sections with the present cosmic-ray energy spectrum provides a production ratio $N_{{}^{10}\text{B}+{}^{11}\text{B}}/N_{{}^9\text{Be}} \approx 14$ for a solar relative distribution of the target elements¹⁷. The smallest production ratio for a broad-band cosmic-ray energy distribution²⁵ is $N_{{}^{10}\text{B}+{}^{11}\text{B}}/N_{{}^9\text{Be}} \approx 8$ for a relative distribution of the elements like that of our metal-poor stars.

The production of ${}^7\text{Li}$ occurs by two pathways. In addition to spallation production from protons on oxygen, the process that dominates boron and beryllium production, there is additional production from ${}^4\text{He}$ nuclei targets, which are always abundant, because the primordial helium abundance is so high. Allowing for this dependence, Steigman and Walker¹⁷ show that the ${}^7\text{Li}$ production will follow $\log(N_{{}^7\text{Li}}/N_{\text{Be}}) \approx -[\text{Fe}/\text{H}]$, if the early cosmic-ray energy spectrum was like that observed now. Hence we expect spallation with this cosmic-ray spectrum to provide a production ratio $(N_{{}^7\text{Li}}/N_{\text{Be}}) \approx 200$ at $[\text{Fe}/\text{H}] = -2.3$. This predicted ratio is more model-dependent than the other spallation constraints. For example, it requires the spallative production of Li from α -particles to occur in the same gas as the proton spallation of Be and B from O, before that gas is consumed in star formation. Thus, assumptions on the efficiency and timescales of mixing of newly enriched material, the lifetimes of cosmic rays, and the timescale and location of star formation are unavoidable. Conversely, one may use these present results to constrain the physical conditions in the early Galaxy. This will be discussed further elsewhere.

We note for completeness that an alternative creation process for beryllium has been proposed—photodissociation of boron²⁶. Besides requiring extreme conditions (carbon-rich gas must be kept within 1 pc of an active galactic nucleus for ~ 1 Gyr before being distributed into the halo, before any metal-poor star formation) this process necessarily produces a high B/Be ratio and is therefore less probable as an explanation of the present observations, where a low B/Be ratio is observed.

The Li observational constraints are shown in Fig. 3. The important feature is that the abundance is independent of $[\text{Fe}/\text{H}]$. The 1σ least-squares upper limit on the slope of the plot of $[\text{Li}/\text{H}]$ against $[\text{Fe}/\text{H}]$ relation for $[\text{Fe}/\text{H}] \leq -1.3$ is 0.05, leading to an upper limit on cosmic-ray production of Li of 1.5×10^{-11} for each unit change in $[\text{Fe}/\text{H}]$. Over the abundance range from $[\text{Fe}/\text{H}] = -3.0$ to $[\text{Fe}/\text{H}] = -2.0$, we derive from the least-squares fits an increase in the Be abundance of 6×10^{-13} . (No useful limits on the Li production are available yet outside this range.) Thus, the $1 - \sigma$ upper limit on the spallative production is $(N_{{}^7\text{Li}}/N_{\text{Be}})_{\text{spall,observed}} \leq 25$, which is one order of magnitude smaller than that reported by Steigman and Walker¹⁷.

A different constraint follows from the B abundance data. Boron abundances for HD140283 ($[\text{Fe}/\text{H}] = -2.8$), HD19445

($[\text{Fe}/\text{H}] = -2.1$) and HD201891 ($[\text{Fe}/\text{H}] = -1.1$) have been reported from Hubble Space Telescope observations (D. Duncan, D. L. Lambert and M. Lemke, manuscript in preparation). They report abundances of $\log(N_{\text{B}}/N_{\text{H}}) = -12.2$, -11.6 and -10.5 , respectively, and note that these are likely to be upper limits. For HD140283, $N_{\text{B}}/N_{\text{Be}} = 5^{+4}_{-2}$. For the other two stars no Be abundance is available, but if they follow the same relationship of $[\text{Be}/\text{H}]$ against $[\text{Fe}/\text{H}]$ as the other metal-poor stars, the expected ratio $N_{\text{B}}/N_{\text{Be}} \approx 10$ for both. These data suggest a lower ratio of B to Be in metal-poor stars than the values of ~ 15 seen in metal-rich stars¹⁶.

These Li/Be and B/Be constraints violate cosmic-ray production limits, assuming a cosmic-ray energy spectrum similar to the present spectrum. The discrepancy is reduced for Li and removed for B if the cosmic-ray energy distribution in the early Galaxy is biased to much higher energies than the present distribution. A further possibility is that local spallation of oxygen in the enriched gas close to supernovae would produce a higher ratio of Be to Li than would spallation in the homogeneous interstellar medium, implicitly assumed by Steigman and Walker. That is, if cosmic-ray spallation predominantly takes place in (temporarily) metal-rich regions the Li discrepancy can be reduced. A final possibility to explain the boron constraint would be to have the primordial Be abundance near the abundance observed below $[\text{Fe}/\text{H}] = -2.5$. Of course, all three explanations might be correct at the same time.

TABLE 2 Iron, oxygen and beryllium abundances

HR	HD	[Fe/H]	Ref.	$\log(N_{\text{Be}}/N_{\text{H}})$	Ref.	[O/H]	Ref.
458	9826	0.04	*	-11.00	29	0.03	§
646	13555	-0.32	*	-11.06	29	-0.21	§
	16031	-1.96	†	-12.37	†	-1.32	†
799	16895	-0.02	*	-10.54	29	-0.01	§
818	17206	0.00	‡	-10.90	29	0.00	§
937	19373	0.07	‡	-10.88	29	0.05	§
1101	22484	-0.11	*	-10.75	29	0.09	§
1173	23754	0.09	*	-11.00	29	-0.06	§
1292	26462	-0.01	‡	-10.86	29	-0.01	§
1543	30652	-0.03	‡	-10.90	29	-0.02	§
	34328	-1.87	†	-11.9	†	-1.46	†
1983	38393	-0.07	*	-10.95	29	-0.14	*
2047	39587	-0.02	*	-11.03	29	-0.02	§
3018	63077	-0.78	*	-11.20	†	-0.44	†
3262	69897	-0.26	*	-10.67	29	-0.17	§
3578	76932	-0.82	*	-11.30	14	-0.53	*
3775	82328	-0.20	*	-10.65	29	-0.13	§
4540	102870	0.13	*	-11.12	29	0.05	*
4785	109358	-0.19	*	-10.58	29	-0.12	§
4825	110379	-0.06	‡	-11.00	29	-0.04	§
4826	110380	-0.06	‡	-11.12	29	-0.04	§
4983	114710	0.03	*	-10.58	29	0.02	§
	116064	-2.18	‡	-12.75	†	-1.74	†
5447	128167	-0.41	*	-11.56	29	-0.27	§
	134169	-1.02	30	-11.35	14	-0.66	§
	140283	-2.77	†	-12.97	†	-2.10	†
5868	141004	-0.04	*	-10.90	29	-0.03	§
5914	142373	-0.52	*	-10.60	29	-0.34	§
5933	142860	-0.16	*	-10.95	29	-0.10	§
	160617	-2.06	†	-12.47	†	-1.68	†
7061	173667	-0.11	*	-11.62	29	-0.07	§
7534	187013	-0.13	*	-10.86	29	-0.08	§
	213657	-2.28	†	-12.65	†	-1.65	†
8665	215648	-0.32	*	-10.56	29	-0.21	§
8969	222368	-0.17	*	-10.79	29	-0.11	§
8181	203608	-0.67	*	-11.40	†	-0.53	†
8430	210027	0.04	‡	-10.67	29	0.03	§
8447	210302	0.10	‡	-10.82	29	0.07	§
	213657	-2.28	†	-12.65	†	-1.65	†
CD - 30° 18140		-2.11	†	-12.28	†	-1.35	†

Results are for all stars that have been studied at high dispersion, and with positive identifications of Be.

* B.E., J. Andersen, B.G., D. L. Lambert, P.E.N. and J. Tomkin, manuscript in preparation.

† P.E.N., B.E., B.G. and G.G., manuscript in preparation.

‡ $[\text{Fe}/\text{H}]$ from Strömgren photometry with calibrations from B.E., J. Andersen, B.G., D. L. Lambert, P.E.N. and J. Tomkin, manuscript in preparation.

§ $[\text{O}/\text{H}] = 0.65 \times [\text{Fe}/\text{H}]$, from B.E., J. Andersen, B.G., D. L. Lambert, P.E.N. and J. Tomkin, manuscript in preparation.

We note also that our Be data, if assumed to reflect a primordial abundance, would not be easily made consistent with standard Big Bang models with a high primordial Li abundance²⁷. In such models the Li has to be removed by high destruction losses, which would however also deplete the Be abundance by a factor up to ~ 5 , requiring the early Be abundance to exceed the standard model prediction by nearly five orders of magnitude.

We propose that the combination of a different cosmic-ray energy spectrum in the early Galaxy and the correlated production of targets and cosmic rays with resulting local spallation explain our data. We cannot exclude the possibility that 50% of the observed Be in the most metal-poor dwarfs observed so far might have been produced in an early process that did not create significant amounts of Li, but did create B with a B/Be production ratio of order unity. The presumably primordial Be/H production value would then exceed the prediction of homogeneous standard Big Bang models^{6,15} by between two and four orders of magnitude, and would suggest an inhomogeneous Big Bang. Even with the cosmic-ray spallation model for all the observed Be in the stars reported here, the observed Be abundance at a heavy element abundance three orders of magnitude

below that in the Sun exceeds that predicted in the most recent calculations of the standard homogeneous Big Bang model by nearly five orders of magnitude.

Conclusion

Combining our new measurements of the abundances of Be and O with observations of Li and B gives bounds on conditions in the very early Galaxy, and may also have cosmological implications. It is likely that the cosmic-ray energy spectrum in the very early Galaxy was biased to high energies, much more than 100 MeV per nucleon, and that the observed Be abundance in our most metal-poor star was essentially created by spallation. It is possible (although not required by the data) that the primordial ⁹Be abundance is significantly in excess of the prediction from the homogeneous standard model.

Finally, we should stress that the above discussion relies on the proper identification and analysis of the two Be II lines in about 10 stars and a single B I line traced in three stars. Precise measurements of the B abundance in other metal-poor stars would be valuable, as would measurements of Be in stars still poorer in heavy elements. Observations with these aims are scheduled. □

Received 10 January; accepted 8 April 1992.

1. Peebles, P. J. E., Schramm, D. N., Turner, E. L. & Kron, R. G. *Nature* **352**, 769–776 (1991).
2. Applegate, J., Hogan, C. & Scherrer, R. *Astrophys. J.* **329**, 572–579 (1988).
3. Ellis, J., Kapusta, J. I. & Olive, K. A. *CERN preprint CERN-TH.6210/91* (1991).
4. Fukugita, M. & Hogan, C. J. *Nature* **354**, 17–18 (1991).
5. Terasawa, N. & Sato, K. *Astrophys. J.* **362**, L47–L49 (1990).
6. Schramm, D. N., Fields, B. & Thomas, D. in *Quark Matter 91* (1991).
7. Gilmore, G., Edvardsson, B. & Nissen, P. E. *Astrophys. J.* **378**, 17–21 (1991).
8. Boesgaard, A. M. & Steigman, G. A. *Rev. Astr. Astrophys.* **23**, 319–378 (1985).
9. Rebolo, R., Molaro, P. & Beekman, J. E. *Astr. Astrophys.* **192**, 192–205 (1988).
10. Schuster, W. J. & Nissen, P. E. *Astr. Astrophys.* **222**, 69–81 (1989).
11. Holweber, H., Band, A., Kock, A. & Kock, M. *Astr. Astrophys.* **249**, 545–549 (1991).
12. Bismont, E., Baudoux, M., Kurucz, R., Ausbacher, W. & Pinnington, E. H. *Astr. Astrophys.* **249**, 539–544 (1991).
13. Deliyannis, C. & Pinsonneault, M. *Astrophys. J.* **365**, L67–L71 (1990).
14. Ryan, S. G., Norris, J. E., Bessell, M. S. & Deliyannis, C. *Astrophys. J.* **388**, 184–189 (1992).
15. Kajino, T. & Boyd, R. N. *Astrophys. J.* **359**, 267–276 (1990).
16. Walker, T. P., Mathews, G. J. & Viola, V. E. *Astrophys. J.* **299**, 745–751 (1985).

17. Steigman, G. & Walker, T. P. *Astrophys. J.* **385**, L13–L16 (1992).
18. Reeves, H. *Proc. IAU Symp.* **145**, 3–12 (Kluwer, Dordrecht, 1991).
19. Vangioni-Flam, E., Cassé, M., Audouze, J. & Obereto, Y. *Astrophys. J.* **364**, 568–582 (1990).
20. Reeves, H. *Phys. Rep.* **201**, 335–354 (1991).
21. Mathews, G. J., Alcock, C. R. & Fuller, G. M. *Astrophys. J.* **349**, 449–457 (1990).
22. Steigman, G. *Phys. Rep.* (in the press).
23. Malaney, R. A. *Phys. Rep.* (in the press).
24. Silberberg, R. & Tsao, C. H. *Phys. Rep.* **191**, 351–408 (1990).
25. Walker, T., Steigman, G., Schramm, D., Olive, K. & Fields, B. *Astrophys. J.* (submitted).
26. Boyd, R. N. & Fencil, H. S. *Astrophys. J.* **373**, 84–88 (1991).
27. Mathews, G. J., Alcock, C. R. & Fuller, G. M. *Astrophys. J.* **349**, 449–457 (1990).
28. Spite, F. & Spite, M. *Astr. Astrophys.* **115**, 357–366 (1982); **141**, 56–60 (1984); **163**, 140–144 (1986).
29. Boesgaard, A. M. *Astrophys. J.* **210**, 466–474 (1976).
30. Magain, P. *Astr. Astrophys.* **179**, 176–180 (1987).

ACKNOWLEDGEMENTS. We thank S. Feltzing for discussions and comments on the manuscript. B.E. is supported by the Swedish Natural Sciences Research Council.

A chimaeric ribonuclease-inhibitor gene restores fertility to male sterile plants

Celestina Mariani*, Veronique Gossele*, Marc De Beuckeleer*, Marc De Block*, Robert B. Goldberg†, Willy De Greef*‡ & Jan Leemans*

* Plant Genetic Systems NV, J. Plateastraat 22, B-9000 Gent, Belgium

† Department of Biology, University of California, Los Angeles, California 90024-1606, USA

Male fertility was restored to genetically engineered male sterile oilseed rape plants. Male sterile plants that express a chimaeric ribonuclease gene in the anther tapetal cell layer were crossed with male fertile plants that were transformed with a chimaeric tapetal-cell-specific ribonuclease-inhibitor gene. F₁ progeny expressing both genes are restored to male fertility by the suppression of cytotoxic ribonuclease activity in the anther by the formation of cell-specific RNase/RNase inhibitor complexes. Genetically engineered male sterility and fertility restorer genes should facilitate hybrid seed production in crop plants.

THE improvement of crop plants through the production of hybrid varieties is a major goal of plant breeding¹. Crosses between inbred plant lines often result in progeny with higher yield, increased resistance to disease, and enhanced perform-

ance in different environments compared with the parental lines¹. The molecular basis of this hybrid vigour is not understood.

The production of hybrid seed on a large scale is challenging because many crops have both male and female reproductive organs (stamen and pistil) on the same plant, either within a single flower (for example oilseed rape, tomato) or in separate

‡ Present address: Laboratory of Genetics, University of Gent, B-9000, Gent, Belgium.

flowers (for example corn). This arrangement results in a high level of self-pollination and makes large-scale directed crosses between inbred lines difficult to accomplish. To guarantee that outcrossing will occur to produce hybrid seed, breeders have either manually or mechanically removed stamens from one parental line, used natural self-incompatibility systems that prevent self-pollination, or exploited male sterility mutations that disrupt pollen development¹⁻⁴. Each of these strategies presents its own set of problems. Manual emasculatation is labour intensive and impractical for plants with small bisexual flowers (such as oilseed rape), many crop plants do not have self-incompatibility and/or male sterility genes, and use of male sterility requires a fertility restorer system⁴.

Recently, we established a genetic engineering strategy for hybrid seed production by demonstrating that chimaeric RNase T1 and barnase genes containing the tobacco *TA29* gene promoter^{5,6} can induce male sterility in tobacco and oilseed rape plants⁷. The *TA29* gene is highly regulated and is transcribed specifically in tapetal cells that surround the pollen sacs in the anther^{6,7}. Expression of the cytotoxic *TA29*-RNase genes selectively destroys the tapetal cell layer, prevents pollen formation, and results in male sterility⁷. The *TA29*-barnase gene contains the coding sequence for the extracellular RNase of *Bacillus amyloliquefaciens*⁷⁻⁹, which has a corresponding inhibitor protein, called barstar^{8,9}. Barstar is produced intracellularly and protects the bacteria from the lethal effects of barnase by forming a stable complex with barnase in the cytoplasm^{8,9}.

Here we show that crosses between male fertile oilseed rape plants containing a *TA29*-barstar gene and male sterile plants containing a *TA29*-barnase gene produce progeny with both genes that are male fertile. The *TA29*-barstar and *TA29*-barnase genes are co-expressed in anthers of the male fertile progeny, indicating that the *TA29*-barstar gene is a dominant suppressor of cytotoxic *TA29*-barnase gene activity, and that fertility restoration is due to the formation of tapetal-cell-specific barnase/barstar complexes. The availability of these genetically engineered nuclear male sterility and fertility restorer genes should facilitate the development of new breeding and production systems for hybrid crops.

Chimaeric *TA29*-barstar expression

We fused the barstar gene coding sequence⁹ with a 1.5-kilobase (kb) *TA29* gene upstream fragment that contains all regulatory elements necessary for tapetal-specific expression⁵⁻⁷. We introduced the *TA29*-barstar gene into oilseed rape plants by Ti-plasmid-mediated transformation using the *bar* gene (bialaphos resistance) as a selectable marker^{10,11} and regenerated 41 transformants that contained 1-5 copies of the chimaeric gene. The *TA29*-barstar transformants were male fertile, produced normal flowers (Fig. 1a), and had well developed anther tapetal cell layers (Fig. 1b). Anther messenger RNA gel blots from several different transformants demonstrated the presence of a prevalent 0.5-kb barstar mRNA, indicating that the *TA29*-barstar gene was expressed correctly (data not shown).

Plants with both genes are male fertile

To determine whether *TA29*-barstar gene expression could inhibit barnase activity in tapetal cells and lead to male fertility restoration, we selected four single-copy *TA29*-barstar gene transformants and used these as male parents in crosses with four male sterile *TA29*-barnase single-copy gene lines⁷ (Fig. 2a). If co-expression of these genes in the anther leads to the formation of stable barnase-barstar complexes, then there should be a 2:1 ratio of fertile to sterile plants in the F_1 progeny after removal of bialaphos-sensitive segregants (Fig. 2b). By contrast, if *TA29*-barstar gene expression does not lead to male fertility restoration, there should be a 1:2 F_1 ratio of fertile to sterile plants (Fig. 2b).

All crosses produced F_1 progeny that segregated with the 3:1 bialaphos-resistant (*HR*) to bialaphos-sensitive (*hr*) ratio ex-

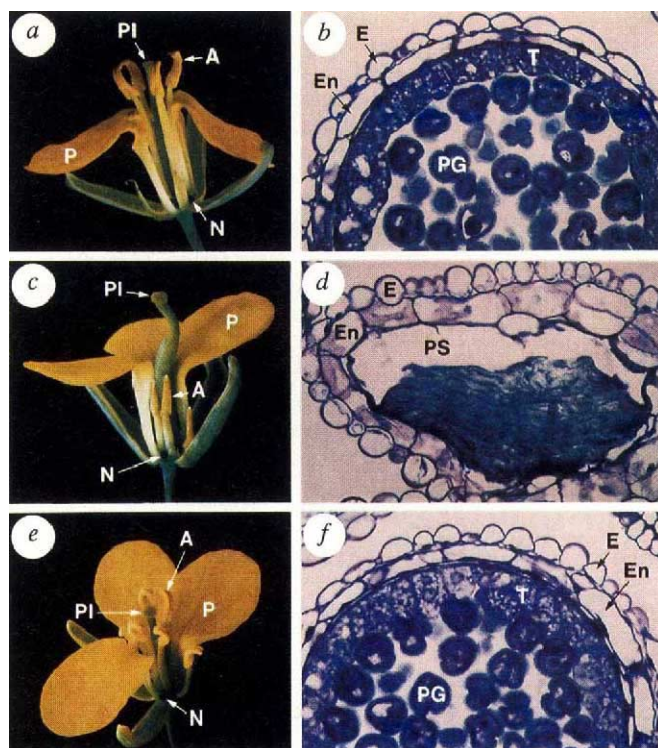


FIG. 1 Oilseed rape flowers and anther cross-sections. a, c and e, Flowers from male-fertile plants containing the *TA29*-barstar gene (a), male-sterile plants containing the *TA29*-barnase gene (c), and male-fertile plants restored to fertility containing both the *TA29*-barstar and *TA29*-barnase genes (e). A, P, Pl and N refer to anther, petal, pistil and nectary, respectively. b, d, f, Bright-field photographs of anther cross-sections from male-fertile plants containing the *TA29*-barstar gene (b), male-sterile plants containing the *TA29*-barnase gene (d), and plants restored to male fertility that contain both the *TA29*-barnase and *TA29*-barstar genes (f). E, En, PG, PS and T refer to epidermis, endothecium, pollen grain, pollen sac and tapetum, respectively.

METHODS. The *TA29*-barstar and *TA29*-barnase genes were placed independently in an *Agrobacterium* vector system containing the *neo* and *bar* genes as selectable markers^{7,10} and used to transform oilseed rape (*Brassica napus* cv. 'Drakkar') hypocotyls by the procedure of De Block *et al.*¹¹. Anthers from 4-mm flower buds were collected, embedded in Histo-resin, sliced into 1-2 μ m sections and stained with 0.05% toluidine blue¹².

pected for parents that contained a single copy of each chimaeric gene (Fig. 2b). The genotypes of the bialaphos-resistant survivors were checked using the polymerase chain reaction (PCR) and male fertility phenotypes were scored by examining pollen release at anther dehiscence (Fig. 2b, c). Six of the nine crosses produced F_1 progeny consistent with the expected 2:1 ratio of male-fertile to male-sterile plants, indicating that male fertility restoration had occurred (shaded boxes, Fig. 2c). All F_1 progeny from these crosses that contained both chimaeric genes were male fertile (numbers in parentheses, Fig. 2c), indicating that the *TA29*-barstar gene functioned as a dominant suppressor of *TA29*-barnase gene activity. Selfing these *TA29*-barstar/*TA29*-barnase plants (progeny of 88-18 $\times$ 94-10 cross; Fig. 2b, c) and removing bialaphos-sensitive segregants produced a 12:3 F_2 ratio of male-fertile to male-sterile plants, indicating that male fertility restoration was co-inherited with the *TA29*-barstar gene and was passed on to the next generation (data not shown).

Two of the initial crosses (88-11 $\times$ 93-101 and 88-11 $\times$ 94-3; Fig. 2b, c) failed to show fertility restoration and produced a 1:2 F_1 ratio of male fertile to male sterile plants (Fig. 2c). PCR analysis of the fertile progeny in these crosses indicated that they were segregants that contained only the *TA29*-barstar gene (Fig. 2b, c). None of the plants containing both chimaeric genes were male fertile (Fig. 2c). In addition, one cross (91-4 $\times$ 94-10; Fig. 2c) produced an intermediate male fertile to male sterile

F₁ ratio (Fig. 2c). We do not know the reason for the absence of full male fertility restoration in these three crosses. However, to eliminate the cytotoxic effects of barnase in the tapetum, the amount of barstar must be equal to or greater than that of barnase⁸. Thus it is likely that the *TA29-barstar* gene in the non-restored plants was expressed at a lower level than the *TA29-barnase* gene.

Restored fertility anthers develop normally

We compared anthers restored to male fertility (MS/RF) with those produced by wild-type plants (WT), male sterile *TA29-barnase* plants (MS), and male fertile *TA29-barstar* plants (RF). Wild-type and *TA29-barstar* plants had anthers that were above the petals (Fig. 1a), contained well developed tapetal cell layers (Fig. 1b), and produced functional pollen grains (Figs 1b and

3a)⁷. Male sterile anthers remained below the petals (Fig. 1c), lacked a tapetum (Fig. 1d), and did not contain viable pollen grains (Figs 1d and 3b)⁷. By contrast, *TA29-barstar/TA29-barnase* anthers restored to male fertility were indistinguishable from those of wild-type and *TA29-barstar* plants (Fig. 1e), had a normal tapetal cell layer (Fig. 1f), and produced large amounts of functional pollen grains that were identical in structure to those produced by either wild-type or *TA29-barstar* plants (Figs 1f and 3c, d, e).

Restored anthers express both chimaeric genes

We analysed anthers of *TA29-barstar/TA29-barnase* plants restored to male fertility (Fig. 2c) for the presence of barstar, barnase and their mRNAs. The RNA gel blots shown in Fig. 4 demonstrate that both barnase and barstar mRNAs were present in anthers of one plant restored to fertility (MS/RF) by the *TA29-barstar* gene. Similar results were obtained with six independent *TA29-barstar/TA29-barnase* plants restored to fertility (data not shown). By contrast, neither of these mRNAs were detected in wild-type or male sterile anthers (Fig. 4). The absence of detectable barnase mRNA in male sterile anthers is due to tapetal cell RNA hydrolysis by barnase activity⁷.

A two-dimensional gel of wild-type anther proteins shows the presence of at least 100 distinct protein spots (Fig. 5a). Proteins from *TA29-barnase* anthers, by contrast, were much fewer in number, suggesting that the missing spots represented tapetal cell and/or pollen grain proteins that were absent in male sterile anthers (Fig. 5b). The two-dimensional protein pattern from *TA29-barstar/TA29-barnase* anthers restored to fertility (MS/RF), however, was indistinguishable from that obtained

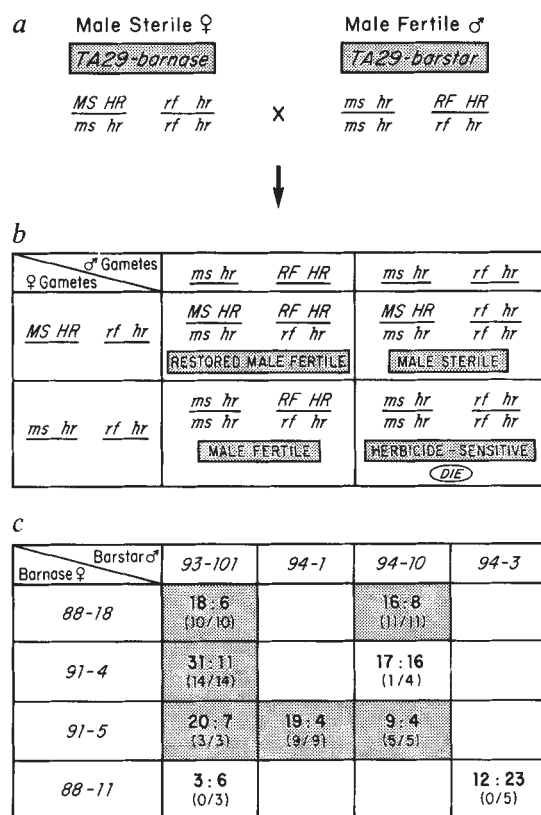


FIG. 2 Restoration of male fertility by crossing oilseed rape plants containing the *TA29-barstar* and *TA29-barnase* genes. **a**, Genotypes of male sterile plants (*TA29-barnase*) and male fertile restorer plants (*TA29-barstar*). *MS*, *RF* and *HR* designate the *TA29-barnase*, *TA29-barstar*, and *bar* herbicide-resistance genes, respectively. *ms*, *rf* and *hr* refer to hemizygous chromosomal loci that lack the *TA29-barnase*, *TA29-barstar* and *bar* genes, respectively. **b**, Genotypes and phenotypes of progeny expected from a cross between male sterile *TA29-barnase* and male fertile restorer *TA29-barstar* lines containing linked herbicide-resistance genes. **c**, Progeny obtained from a cross between male sterile *TA29-barnase* and male fertile *TA29-barstar* plants. Ratio in each box refers to the actual number of male fertile to male sterile plants scored in the progeny. Number in parenthesis below each ratio refers to the number of progeny containing both the *TA29-barnase* and the *TA29-barstar* genes that were male fertile. Boxes shaded in grey highlight crosses in which the *TA29-barstar* gene suppressed *TA29-barnase* gene activity and restored fertility to male sterile plants. **METHODS**. Oilseed rape plants containing a single-copy of either the *TA29-barnase*⁷ or the *TA29-barstar* genes were identified by DNA gel blot analysis. These plants were crossed as diagrammed in **a** and seeds were collected from mature siliques. Seeds from each cross were planted and young seedlings were sprayed with bialaphos to kill herbicide-sensitive plants that did not contain either the *TA29-barnase* or the *TA29-barstar* genes as shown in **b**. The genotypes of herbicide-resistant progeny were determined by PCR analysis and the phenotypes were scored by the presence or absence of viable pollen on the anthers at dehiscence.

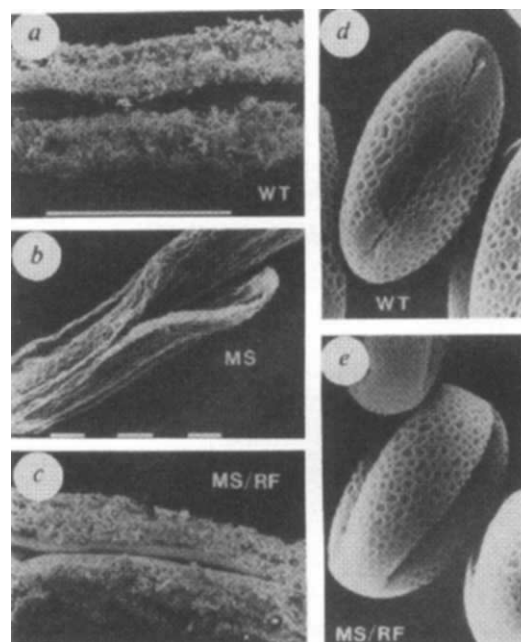


FIG. 3 Scanning electron micrographs of oilseed rape pollen grains and dehiscing anthers. **a**, Dehiscing anthers from an untransformed plant. Scale bar, 1 mm; magnification factor, ×69. **b**, Dehiscing male sterile anther from a plant containing the *TA29-barnase* gene. Scale bar, 0.1 mm; magnification factor, ×131. **c**, Dehiscing anther from a plant restored to male fertility containing both the *TA29-barstar* and *TA29-barnase* genes. Scale bar, 1 mm; magnification factor, ×69. **d**, **e**, Pollen grains from untransformed anthers (**d**), and anthers containing both the *TA29-barstar* and *TA29-barnase* genes that were restored to fertility (**e**). Scale bars, 10 μm; magnification factors, ×2,400.

METHODS. Pollen grains and dehiscing anthers were collected from open oilseed rape flowers and photographed by scanning electron microscopy as outlined previously⁷. Dehiscing anthers were removed from 7-mm flower buds and fixed before scanning electron microscopy as described^{6,7}.

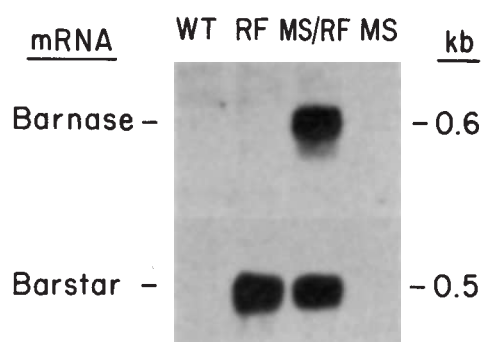


FIG. 4 Presence of barstar and barnase mRNAs in anthers of oilseed rape plants restored to male fertility. Gel blots containing 1 μ g anther poly(A)⁺ mRNAs from untransformed plants (WT), *TA29*-barstar plants (RF), male fertile plants containing both the *TA29*-barnase and *TA29*-barstar genes (MS/RF), and male sterile *TA29*-barnase plants (MS) were hybridized sequentially with *TA29*-barnase and *TA29*-barstar gene probes.

METHODS. Anthers were collected from 4-mm flower buds, total RNAs were isolated as described¹³, and poly(A)⁺ mRNAs were selected according to the protocol supplied with the Pharmacia-LKB poly(A)⁺ RNA purification kit.

with wild-type anther proteins (Fig. 5a, c). Immunoblots of these proteins using antibodies raised against the barnase-barstar complex demonstrated the presence both of barnase and of barstar (Fig. 5d, e), indicating that fertility restoration was due to the inhibition of barnase activity by the formation of tapetal-cell-specific barnase-barstar protein complexes.

Discussion

We have shown previously that the *TA29*-barnase gene acts as a dominant male sterility gene both in tobacco and in oilseed rape plants⁷. *TA29*-barnase gene expression leads to the selective destruction of tapetal cells during anther development and prevents the formation of functional pollen grains⁷. The results presented here show that the *TA29*-barstar gene suppresses *TA29*-barnase gene expression in oilseed rape anthers by protein-protein interactions, protects tapetal cells from barnase cytotoxic activity, and restores male fertility. The *TA29*-barstar gene also acts as a dominant restorer of male fertility in tobacco, indicating that barstar is able to work effectively in different plants (C.M. *et al.*, unpublished results).

Barnase and barstar are single-chain proteins that can function under a variety of conditions⁸. Barnase inhibition is due to the formation of a diffusion-dependent, one-to-one complex between barnase and barstar⁸. This complex is extremely stable and has a dissociation constant of about 10^{-14} M, indicating that once it forms it rarely dissociates. The *TA29*-barstar and *TA29*-barnase genes are both controlled by the same tapetal-specific regulatory sequences^{6,7}. We adopted this strategy to ensure that each gene would be activated in tapetal cells at the same time and to maximize the chance that barstar molecules would accumulate in equal or greater amounts than barnase in the tapetal cell layers.

The majority of crosses between *TA29*-barstar and *TA29*-barnase plants produced progeny with both chimaeric genes that are male fertile (Fig. 2c), indicating that this strategy was successful. Anthers restored to male fertility are indistinguishable from those of wild-type plants, develop and dehiscence normally (Figs 1 and 3), have well-differentiated tapetal cell layers (Fig. 1f), and produce large amounts of functional pollen grains (Figs 1 and 3). These results show that barstar is able to complex efficiently with barnase in anther tapetal cells and that use of the *TA29*-barstar gene as a dominant restorer of male fertility only requires that it is expressed equally or to a greater extent than the *TA29*-barnase gene.

In addition to oilseed rape⁷, we have shown that the tobacco *TA29* gene promoter is active in other crop plants, including lettuce, chicory, cauliflower, tomato, cotton and corn (A. Reynaerts, K. D'Halluin and C.M., unpublished results). In

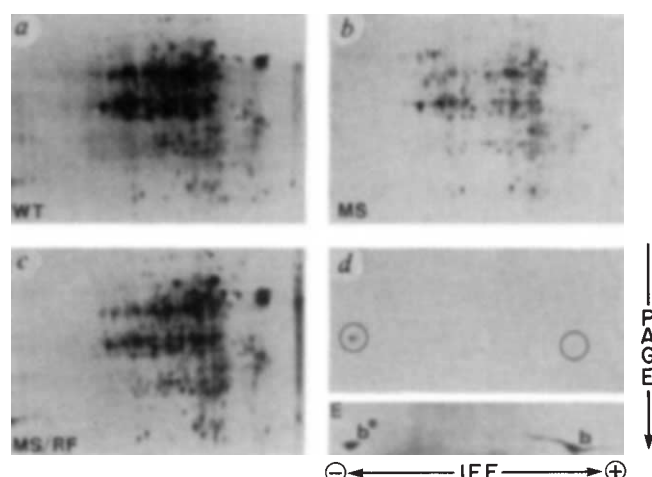


FIG. 5 Presence of barstar and barnase proteins in oilseed rape anthers restored to male fertility. a, b, c, d, Anther proteins from wild-type (WT), untransformed plants (a), male sterile (MS) plants containing the *TA29*-barnase gene (b), and plants restored to male fertility (MS/RF) that contain both the *TA29*-barnase and *TA29*-barstar genes (c, d) were fractionated by two-dimensional gel electrophoresis and either stained with silver (a, b, c) or blotted and reacted with antibodies specific for the barnase-barstar complex (d). Barstar and barnase proteins present in anthers restored to male fertility are circled in the immunoblot (d). e, Purified barnase-barstar complexes (1 μ g) were denatured, fractionated by two-dimensional gel electrophoresis, and stained with silver. b* and b refer to barstar and barnase, respectively.

METHODS. Oilseed rape anthers were collected from 4-mm flower buds and proteins were extracted by grinding anthers (50 mg) in 500 μ l 50 mM phosphate buffer (pH 7), 10 mM EDTA, 10 mM β -mercaptoethanol, 0.1% Triton X-100, 0.1% sarcosyl, 0.6% polyvinylpyrrolidone, and 25 μ g ml⁻¹ PMSF. Proteins were precipitated with 10% TCA, washed with 90% acetone, dried, and resuspended in Pharmacia lysis buffer according to the protocol supplied by Pharmacia-LKB. Protein fractionation by two-dimensional gel electrophoresis was done using the Pharmacia-LKB Multiphor II 2-D System following the protocol supplied by the manufacturer. Total anther protein (5 μ g) was used for the silver-stained gels (a, b, c) and 200 μ g protein for the immunoblot (d). The immunoblot analysis was done according to the Pharmacia-LKB protocol.

each crop, the expression of the *TA29*-barnase gene leads to the production of male sterile plants (A. Reynaerts, K. D'Halluin and C.M., unpublished results). Use of the *TA29*-barnase gene alone in plants such as lettuce and chicory should permit the efficient production of hybrid plants because leaves are the harvested product. By contrast, both the *TA29*-barstar and *TA29*-barnase genes will be required to produce hybrid seeds and fruit in plants such as tomato and oilseed rape. The effectiveness of the chimaeric *TA29*-barnase and *TA29*-barstar genes in male sterility induction and male fertility restoration should permit the breeding of genetically engineered hybrid crop plants in the near future.

Received 26 February; accepted 10 April 1992.

- Mayo, H. *The Theory of Plant Breeding* (Clarendon, Oxford, 1980).
- Haring, V. *et al. Science* **250**, 937-941 (1990).
- Levings, C. S. *Science* **250**, 942-947 (1990).
- Kaul, M. L. H. *Monographs on Theoretical and Applied Genetics* Vol. 10 *Male Sterility in Higher Plants* (Springer, Berlin, 1988).
- Seurinck, J., Truettner, J. & Goldberg, R. B. *Nucleic Acids Res.* **18**, 3403 (1990).
- Koltunow, A. M. *et al. Plant Cell* **2**, 1201-1224 (1990).
- Mariani, C. *et al. Nature* **347**, 737-741 (1990).
- Hartley, R. W. *Trends Biochem. Sci.* **14**, 450-454 (1989).
- Hartley, R. W. *J. molec. Biol.* **202**, 913-915 (1988).
- De Block, M. *et al. EMBO J.* **6**, 2513-2518 (1987).
- De Block, M., De Brouwer, D. & Tenning, P. *Pl. Physiol.* **91**, 695-701 (1989).
- De Block, M. & De Brouwer, D. *Pl. J.* **2**, 261-266 (1992).
- Jones, D. G., Dunsmuir, P. & Bedbrook, J. *EMBO J.* **4**, 2411-2418 (1985).

ACKNOWLEDGEMENTS. We thank R. Hartley for the barstar gene and the antibodies against the barnase-barstar complex; M. Lauwereys, P. Stanssens, J. Steyaert and M. Bielefeld for sharing their knowledge of RNases; D. De Brouwer and B. Van Steenbrugge for technical help; S. Jansens for photography; and M. Cornelissen, W. Timberlake and R. Chasen for reading this manuscript. Part of this work was supported by the NSF (R.B.G.).

Can accretion onto isolated neutron stars produce γ -ray bursts?

Alice K. Harding & Marvin Leventhal

Laboratory for High Energy Astrophysics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

THE isotropy and flat count spectrum of γ -ray bursts revealed by the BATSE detector on the Compton Gamma-Ray Observatory¹ have led to suggestions that the burst sources are an extended galactic halo of high-velocity neutron stars^{2,3}. We show here that if slow accretion onto these neutron stars from the interstellar medium is to be the origin of γ -ray bursts, the accretion physics is very different from what applies for local, low-velocity neutron stars^{4,5}. For halo neutron stars with high magnetic fields and velocities ($v > 190 \text{ km s}^{-1}$), electromagnetic dipole radiation pressure prevents accretion unless the period is longer than tens of seconds; the centrifugal barrier will then prevent accretion until the period reaches several thousand seconds. For periods as long as this, accretion may proceed through Kelvin–Helmholtz instability at the magnetopause boundary. At interstellar densities and neutron-star magnetic fields of $\sim 10^{12} \text{ G}$, the accretion rate by this process can be much larger than the Bondi–Hoyle (hydrodynamic) accretion rate, but is still well below what is needed for slow-accretion burst models. We conclude that slow accretion onto high-velocity neutron stars in the halo cannot be the origin of γ -ray bursts.

Thermonuclear flash models for γ -ray bursts require accretion rates $\dot{M} \approx 10^{-14} - 10^{-16} M_{\odot} \text{ yr}^{-1}$, where $M_{\odot}$ is the solar mass, for slow accumulation of the $10^{21} - 10^{23} \text{ g}$ needed to trigger the helium flash through hydrogen burning^{4,6}. If the effect of the neutron star magnetic field is neglected, the interstellar accretion rate is given by the Bondi–Hoyle accretion formula, $\dot{M}_{\text{BH}} = \pi n m_p R_s^2 c_s^4 (v^2 + c_s^2)^{-3/2}$, where $R_s = 2GM/c^2$, M and v are the neutron star mass and velocity, c_s is the sound velocity, n is the interstellar density and m_p is the proton mass. Neutron stars near the galactic plane with relatively low velocities ($v < 30 \text{ km s}^{-1}$) can accrete at rates high enough to cause a thermonuclear flash^{4,7}. But long rotation periods are required to overcome the centrifugal barrier, before accretion can take place^{8–10}.

The BATSE results on the burst spatial distribution have now ruled out a local disk population of neutron stars as all but a small fraction of the γ -ray burst sources. Galactic neutron stars are still acceptable candidates, but they must occupy an extended halo of radius $R > 50 \text{ kpc}$, requiring high space velocities^{2,3}. There is in fact evidence from studies of radio pulsars that such a population of neutron stars exists. Proper-motion measurements¹¹ have recently discovered pulsars with velocities up to $1,000 \text{ km s}^{-1}$, some of which are moving towards the galactic plane and many of which are not bound to the Galaxy. Statistical studies¹² suggest two subpopulations of pulsars, one born with average space velocity $\langle v \rangle \approx 90 \text{ km s}^{-1}$, short periods and at low galactic latitude, the other born with $\langle v \rangle \approx 300 \text{ km s}^{-1}$, longer periods and at higher latitude. Silk²³ has suggested that a separate population of high-velocity neutron stars may have formed in the halo by white-dwarf mergers. The number of neutron stars formed by this process could be much higher than that in the disk without violating nucleosynthesis limits.

The accretion scenarios studied for the low-velocity neutron stars in the galactic disk neglect the magnetic field of the star, assuming that the Bondi–Hoyle accretion radius

$$R_A = 2GM/v^2 \quad (1)$$

is larger than the magnetopause radius, R_M , defined by $B_M^2/8\pi = nm_p v^2/2$, where B_M is the value of the dipole field at

R_M , or

$$R_M = \left(\frac{\mu^2}{4\pi n m_p v^2} \right)^{1/6} = 1.3 \times 10^{11} \text{ cm } \mu_{30}^{1/3} n^{-1/6} v_8^{-1/3} \quad (2)$$

where $\mu_{30} \equiv \mu/10^{30} \text{ G cm}^3$ is the neutron star magnetic moment, n is the interstellar density in units of 1 cm^{-3} and $v_8 \equiv v/10^8 \text{ cm s}^{-1}$. This is not a valid assumption for high velocity neutron stars, for which $R_M > R_A$ when

$$v > 380 \text{ km s}^{-1} \mu_{30}^{-1/5} n^{1/10} \left(\frac{M}{M_{\odot}} \right)^{3/5} \quad (3)$$

In this case, ionized gas will be influenced by the magnetic field which dominates the accretion process. A halo population of neutron stars would spend most of their time accreting well above the galactic plane, where the gas is predominantly ionized¹³ with a density $n \approx 10^{-3} \text{ cm}^{-3}$. Thus, stars with magnetic fields of 10^{12} G having $v > 190 \text{ km s}^{-1}$ would have $R_M > R_A$.

As is the case for low-velocity neutron stars, short rotation periods will prevent accretion. Studies of the radio pulsar population indicate that at least this class of neutron stars are born with relatively short periods, in the range 10–500 ms. As long as the light cylinder radius, $R_L = c/\Omega$, of a neutron star with rotation frequency Ω is inside its magnetospheric radius, the electromagnetic dipole radiation pressure exceeds the external pressure, and the star cannot accrete material^{14,15}. Stars having $R_M > R_A$ therefore cannot accrete from the interstellar medium until they have spun down to a period

$$P > P_1 = 27 \text{ s } \mu_{30}^{1/3} n^{-1/6} v_8^{-1/3} \quad (4)$$

defined by the condition $R_L > R_M$. The timescale to reach this period from a birth period $P_0 \ll P_1$ through electromagnetic dipole radiation spin-down is

$$\tau_{\text{EM}} = \frac{3IP_1^2 c^3}{8\pi^2 (\mu \sin \alpha)^2} = 2.4 \times 10^{10} \text{ yr } I_{45} \mu_{30}^{-4/3} n^{-1/3} v_8^{-2/3} \quad (5)$$

where $I_{45} \equiv I/10^{45} \text{ g cm}^2$ is the neutron star moment of inertia and α is the angle between magnetic and spin axes. We have assumed that $\sin \alpha \approx 1$. As this time is roughly the age of the Galaxy, accretion onto a strongly magnetized neutron star from the interstellar medium will not take place unless some spin-down mechanism faster than dipole radiation can operate.

If a period P_1 can somehow be reached, when $R_L \geq R_M$, the neutron star will still not accrete material as long as the centrifugal radius, where the centrifugal force equals the gravitational force,

$$R_C = \left(\frac{GM}{\Omega^2} \right)^{1/3} \quad (6)$$

is larger than R_M . When this situation holds, the neutron star transfers energy and angular momentum to the material just outside the magnetopause. If the magnetic pole is aligned with the spin axis, then friction at the magnetopause will transfer heat and angular momentum to the ambient material¹⁶. If, as is more likely, the magnetic and spin axes are not aligned, then the star can also spin down through the propeller effect¹⁷. To overcome the centrifugal barrier, the neutron star must reach a period of

$$P_{11} = \frac{2\pi}{\sqrt{GM}} R_M^{3/2} = 2 \times 10^4 \text{ s } \mu_{30}^{1/2} n^{-1/4} v_8^{-1/2} \left(\frac{M}{M_{\odot}} \right)^{-1/2} \quad (7)$$

defined by $R_C = R_M$. Using the spin-down rate for the propeller effect^{17,18}

$$I \dot{\Omega} = 4\pi R_M^3 v^2 n m_p \frac{v}{R_M \Omega} \quad (8)$$

we find that the timescale for spin-down from period P_1 to period P_{11} , or $\tau_P = \Omega/2\dot{\Omega}$ evaluated at P_1 , is

$$\tau_P = 2.5 \times 10^{12} \text{ yr } I_{45} \mu_{30}^{-4/3} n^{-1/3} v_8^{-5/3} \quad (9)$$

Again, a more rapid spin-down mechanism is required to reach

the period where accretion can begin. Both P_I and P_{II} are considerably longer than the corresponding period limits for the Bondi-Hoyle (low-velocity) regime¹⁰.

If the neutron star spins down to a period greater than P_{II} and $R_M > R_A$, then the accretion of ionized material must occur at a rate different from the Bondi-Hoyle rate because of the influence of the star's magnetic field. In the plane of the Galaxy, only a few per cent of the gas on average is ionized, although radiation from the neutron star itself is enough to ionize the surrounding gas¹⁹, and at high latitudes the gas becomes predominantly ionized. As the star moves through the ambient medium, the ionized gas will encounter a bow shock in front of the magnetopause, deflecting the flow around the object and forming a layer of shocked gas moving with speed $\sim v$ relative to the magnetosphere (see Fig. 1). Such a boundary layer can develop waves which are Kelvin-Helmholtz unstable with growth timescale²⁰

$$\tau_{KH} \approx \frac{(1 + \rho_i/\rho_e)}{kv(\rho_i/\rho_e)^{1/2}} \quad (10)$$

where $l_{KH} = k^{-1}$ is the wavelength of the fastest growing unstable mode, and ρ_i and ρ_e are the gas densities internal and external to the magnetopause. Equation (10) assumes that $\rho_i/\rho_e \geq a^2/c^2 \approx 10^{-5} v_8^2$, where $a = B_M/\sqrt{4\pi\rho_e}$ is the Alfvén velocity at R_M . The velocity at which plasma crosses the magnetospheric boundary is²⁰ $v_{conv} \approx l_{KH}/\tau_{KH}$, or

$$v_{conv} \approx v(\rho_i/\rho_e)^{1/2}(1 + \rho_i/\rho_e)^{-1}\eta_{KH} \quad (11)$$

where η_{KH} is an efficiency factor of order 0.1. We can then estimate the accretion rate of material onto the neutron star as

$$\begin{aligned} \dot{M}_{KH} &\approx 2\pi R_M^2 \rho_e v_{conv} \\ &= 1.8 \times 10^7 \text{ g s}^{-1} \mu_{30}^{2/3} n^{2/3} v_8^{1/3} \eta_{KH} (\rho_i/\rho_e)^{1/2} (1 + \rho_i/\rho_e)^{-1} \end{aligned} \quad (12)$$

where we have assumed that $\rho_e \approx n$, although it is possible that $\rho_e \gg n$ because of compression of the flow in the boundary layer. Unlike the unmagnetized case where $\dot{M}_{BH} \propto v^{-3}$, the accretion rate in this mechanism increases with neutron star velocity. Alcock and Illarionov¹⁹ suggested that a similar rate may be possible for accretion onto magnetic white dwarfs, but they did not discuss instability at the magnetopause boundary.

To get an estimate of the possible range of $\dot{M}_{KH}$ for fast neutron stars in the halo, we can obtain a lower limit on ρ_i/ρ_e

by taking ρ_i greater than or equal to the corotation charge density at the magnetopause boundary

$$\frac{\rho_i}{\rho_e} \geq \frac{\Omega B_M m_p}{2ec} \frac{1}{\rho_e} = 10^{-4} v_8 P^{-1} n^{-1/2} \quad (13)$$

Then taking $\mu_{30} = v_8 = 1$ and $n = 10^{-3}$ in equation (12), and assuming equation (13) with $P = P_{II}$ and $\eta_{KH} = 0.1$ as a lower limit and $v_{conv} = v$ as an upper limit, we have a mass accretion rate range, $3.0 \text{ g s}^{-1} < \dot{M}_{KH} < 1.8 \times 10^5 \text{ g s}^{-1}$. Comparing this to the corresponding Bondi-Hoyle accretion rate for $v_8 = 1$ and $n = 10^{-3}$, $\dot{M}_{BH} \approx 120 \text{ g s}^{-1}$, it is apparent that $\dot{M}_{KH}$ for a highly magnetized, fast neutron star can be much higher, primarily because $R_M \gg R_A$.

Neutron stars in the halo at an average distance of 30 kpc need 10^{41} erg to make a γ -ray burst of average fluence. A thermonuclear flash powered by triple-alpha helium burning with an efficiency $\epsilon \approx 10^{-3}$ requires an accreted mass of $\dot{M}_{acc} = 10^{23} \text{ g}$ to make a burst of this energy. But even the maximum accretion rate, $\dot{M}_{KH} < 1.8 \times 10^5 \text{ g s}^{-1}$, is well below the minimum accretion rate of $\dot{M} \approx 10^{10} \text{ g s}^{-1}$ required for a thermonuclear flash²¹, because the accreted material will never reach the temperature for explosive nuclear burning. In the model of Blaes *et al.*⁵, slow accretion builds up a layer of non-explosively burning material in the neutron star surface to a threshold accreted mass $\sim 10^{26} \text{ g}$, where overturn instability occurs to cause a γ -ray burst. This model does not require a minimum accretion rate, but the time needed to accrete the threshold mass at a rate of 10^5 g s^{-1} is $\sim 3 \times 10^{13} \text{ yr}$. Therefore, neither the Blaes *et al.* model nor the thermonuclear flash model will work to produce γ -ray bursts on high-velocity neutron stars in the halo. We have implicitly assumed that the neutron-star magnetic field does not decay over long timescales. Although the issue of field decay is controversial, both the thermonuclear flash models and observations of cyclotron lines²² favour strongly magnetized neutron stars as sources of γ -ray bursts. But even if the magnetic fields of isolated neutron stars do decay on relatively short (10^6 – 10^7 yr) timescales, the accretion would shift to the Bondi-Hoyle rate, which we have shown is even lower than $\dot{M}_{KH}$ for these high-velocity neutron stars. Even in the best of circumstances then, where spin-down to a period longer than P_{II} takes place rapidly and accretion at a rate $\dot{M}_{KH}$ occurs with 100% efficiency, a fast neutron star cannot accrete at a rate high enough to make a thermonuclear flash even once over the age of the Galaxy. $\square$

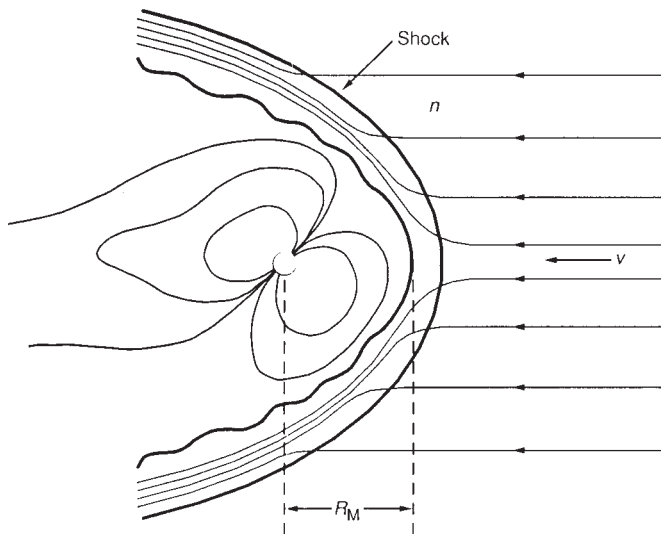


FIG. 1 Schematic view of magnetized neutron star moving through the interstellar medium with speed v . R_M is the magnetopause radius and n is the interstellar density. The distorted dipole magnetic field of the neutron star is indicated and flow lines are shown in the neutron-star rest frame.

Received 13 January; accepted 22 April 1992.

1. Meegan, C. A. *et al.* *Nature* **355**, 143 (1992).
2. Hartmann, D. *Nature* (submitted).
3. Dermer, C. D. & Li, H. in *Proc. Huntsville Gamma-Ray Burst Workshop* (eds Paciesas, W. S. & Fishman, G. J. (in the press)).
4. Hameury, J. M., Bonazzola, S., Heyvaerts, J. & Ventura, J. *Astr. Astrophys.* **111**, 242 (1982).
5. Blaes, O. R., Blandford, R. D., Madau, P. & Koonin, S. *Astrophys. J.* **363**, 612–627 (1990).
6. Woosley, S. E. & Taam, R. E. *Nature* **263**, 101 (1976).
7. Blaes, O. & Rajagopal, M. *Astrophys. J.* **381**, 210–219 (1991).
8. Blaes, O., Blandford, R., Goldreich, P. & Madau, P. *Astrophys. J.* **343**, 839 (1989).
9. Kluzniak, W. & Ruderman, M. in *Proc. of Huntsville Gamma-Ray Burst Workshop* (eds Paciesas, W. S. & Fishman, G. J. (in the press)).
10. Blaes, O., Blandford, R., Madau, P. & Yang, L. *Astrophys. J.* (submitted).
11. Harrison, P. A., Lyne, A. G. & Anderson, B. in *NATO ARW on X-ray Binaries and the Formation of Binary and Millisecond Pulsars* (eds van den Heuvel, E. J. P. & Rappaport, S. J. (Kluwer, Dordrecht, in the press)).
12. Narayan, R. & Ostriker, J. P. *Astrophys. J.* **352**, 222 (1990).
13. Reynolds, R. J. in *The Interstellar Disk-Halo Connection in Galaxies* (ed. Bloemen, H.) 67–76 (Kluwer, Dordrecht, 1991).
14. Ruderman, M. A., Shaham, J. & Tavani, M. *Astrophys. J.* **336**, 507 (1989).
15. Harding, A. K. & Gaisser, T. K. *Astrophys. J.* **358**, 561–574 (1990).
16. Mineshige, S., Rees, M. J. & Fabian, A. C. *Mon. Not. R. astr. Soc.* **251**, 555–563 (1991).
17. Illarionov, A. F. & Sunyaev, R. A. *Astr. Astrophys.* **39**, 185–195 (1975).
18. Davies, R. E., Fabian, A. C. & Pringle, J. E. *Mon. Not. R. astr. Soc.* **186**, 779–782 (1979).
19. Alcock, C. & Illarionov, A. *Astrophys. J.* **235**, 541–553 (1980).
20. Burnard, D. J., Lea, S. M. & Arons, J. *Astrophys. J.* **266**, 174–187 (1983).
21. Lasota, J. P. in *Gamma-Ray Bursts* (eds Ho, C. Epstein, R. I. & Fenimore, E. E.) 17 (Cambridge Univ. Press, 1992).
22. Murakami, T. *et al.* *Nature* **335**, 234 (1988).
23. Silk, J. *Phys. Rep.* (in the press).

ACKNOWLEDGEMENTS. We thank P. Ghosh, D. Hartmann, R. Reynolds, M. Ruderman, B. Savage, B. Schaefer and the referee O. Blaes for comments and advice.

A neutron lens

M. A. Kumakhov & V. A. Sharov

Institute for Roentgen Optical Systems, World-Lab,
I.V. Kurchatov Atomic Energy Institute, Moscow 123182, Russia

SMOOTH hollow glass capillaries can be used to guide neutrons by multiple reflection at grazing angles along the inner walls. Here we show that a concentric arrangement of bent capillaries makes it possible to focus and increase the flux density of a neutron beam—a phenomenon that will have applications in areas such as prompt-gamma activation analysis and neutron depth profiling.

Using a steering device to make the neutron flux more intense and localized improves both spatial resolution and detection limits in applications of thermal and cold neutrons to materials analysis. This also increases the efficiency of neutron capture therapy and reduces radiation exposure. A new technique for X-ray and neutron optics has been proposed¹ which is based on multiple reflection from inner surfaces of curved hollow capillaries. Considerable progress has been made in the creation of X-ray optical systems based on this concept²⁻⁴. We obtained encouraging results from attempts to deflect thermal neutrons from a reactor source in 1989 (ref. 5) and have also measured the single reflection coefficient $R = 0.99$ of thermal neutrons from the inner surface of a lead glass capillary⁶. This is in good agreement with results for ultracold neutrons⁷, where $R = 0.986$, and with the results of another group⁸.

A neutron may be reflected from a smooth surface with a probability close to unity if the incident angle is less than a critical angle, θ_c , which depends on the properties of the material and on the neutron wavelength. Consequently, the neutron can travel along an empty channel or guide while being reflected many times. If the channel has a radius of curvature $R > 2d/\theta_c$ (where d is the diameter of the channel and θ_c the critical angle), the beam propagates along the curved channel. Neutrons from an extended quasiparallel beam (such as from a thermal or cold-neutron guide tube) can be captured by a system of capillaries and directed towards a single focal point as shown in Fig. 1. Here the full cross-section of the incident beam is divided into smaller beams, and each curved channel turns, by an appropriate angle, that part of the radiation which enters the channel within the critical angle θ_c . A channel on the periphery turns the radiation through a larger angle than one near the axis. It is also possible to use as channels a system of nested second-degree surfaces (Fig. 1). Such systems transform the quasiparallel beam into a convergent beam focused at a common spot.

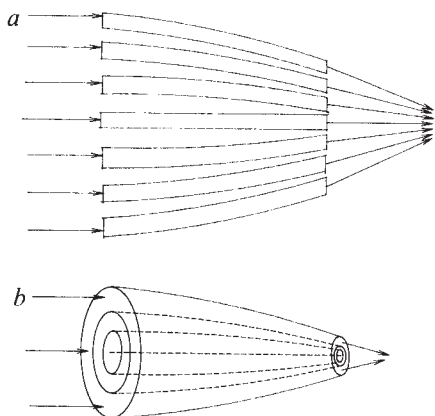


FIG. 1 Schematic diagram of system of (a) hollow capillary tubes and (b) nested two-dimensional surfaces used to transform a quasiparallel neutron beam into a convergent beam.

We have designed a neutron lens (Fig. 2) using a technology developed for X-rays using a lead glass of composition (in mol per cent): SiO_2 , 55.0; PbO , 30.0; Al_2O_3 , 2.0; Na_2O , 3.8; K_2O , 9.2. The lens was assembled from multicapillaries with each fibre containing more than 1,000 parallel channels. Each fibre is hexagonal in cross-section and 0.4 mm in width. The individual channels are 6.0 μm in diameter. A micrograph showing part of the cross-section of a fibre is shown in Fig. 1 of ref. 8.

The lens (Fig. 2) was constructed from 721 fibres, held in place by four thin metal membranes with holes made by chemical

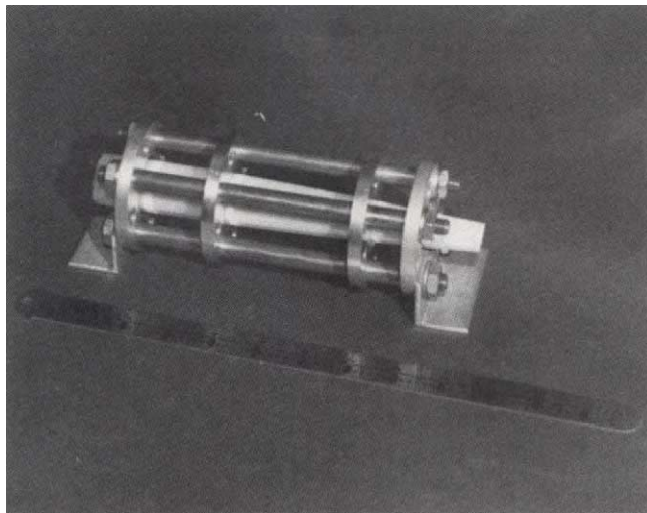


FIG. 2 Photograph of the neutron lens. The lens is 22 cm long; other parameters are described in the text.

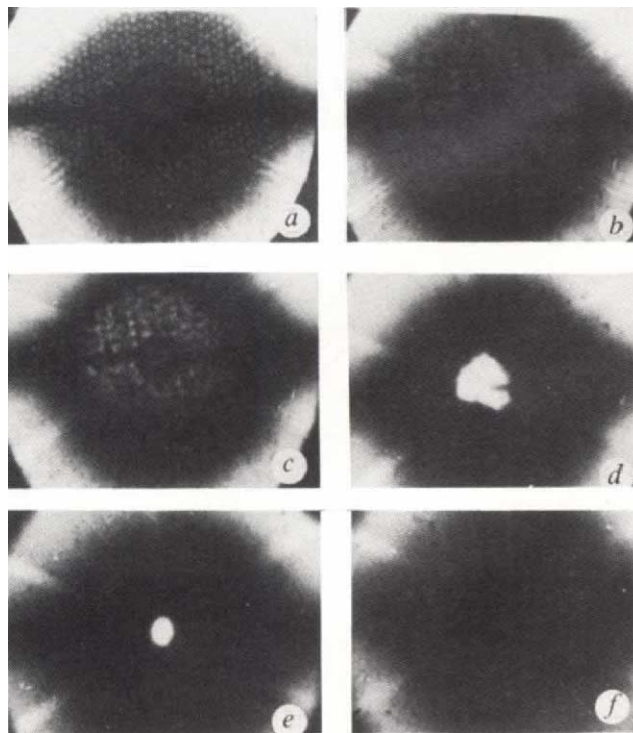


FIG. 3 Photographs of X-ray film and gadolinium foil images made at different distances from the exit of the neutron lens. These images were taken with a cadmium shield over the central region of the lens. The distance from the lens for each image was (a) 0, (b) 20 mm, (c) 50 mm, (d) 80 mm, (e) 100 mm (approximate focus), (f) 200 mm.

etching through a photolithographic mask and spaced so that each fibre follows a calculated path. The radius of bending of the outermost layer of fibres was 1,327 mm. The entrance diameter (large end) of the fibre bundle was 29 mm and the exit diameter was 15 mm. The distance from the fibre bundle to the focal spot was 104 mm, determined by exposing a neutron-sensitive gadolinium foil/film sandwich at closely spaced intervals from the bundle. The neutron divergence at the focal spot was 3.6° around the central axis. All measurements were made on a monochromatic thermal neutron beam at the Kurchatov Institute IR-8 reactor. This lens was aligned with a quasiparallel monochromatic neutron beam with wavelength $\lambda = 2.4 \text{ \AA}$ and divergence $\pm 20 \text{ min of arc}$. The position of the focal spot and the lens efficiency were estimated by exposing a gadolinium foil/X-ray film sandwich at various points along the path of the emergent beam. To avoid possible exposure of the X-ray film by unguided neutrons, we placed a cadmium disc at the entrance of the lens to block the central part of the lens. Figure 3 shows a sequence of cross-sectional photographs depicting the neutron beam pattern and intensity at different distances from the exit of the lens. The output of individual capillaries (Fig. 3a) successively converges to a spot at the focus (Fig. 3e). At twice the focal length (Fig. 3f), one cannot distinguish between individual capillary outputs because of overlap due to divergence of the beam emerging from each capillary channel. The focal distance was determined to be 104 mm and the diameter of the focal spot was 1 mm. By comparing optical densities at the focal spot with

and without the lens, we have found that under these conditions the lens gain is ~ 7 : that is, the lens amplifies the neutron flux density by a factor of 6–7 as compared with the intensity at the same spot without the lens.

The availability of such a neutron-focusing device gives new life to traditional applications in neutron activation analysis and medicine. Usually we are interested in the optimum flux density of the input beam, the beam divergence, focal spot size and maximum flux achieved. Our initial effort has demonstrated a focal spot diameter of 1 mm or slightly smaller. We believe that with extension of this technique we will have the capability to produce a focal spot diameter of $\sim 30 \mu\text{m}$ and a flux density gain of up to 1,000. $\square$

Received 9 January; accepted 8 April 1992.

1. Kumakhov, M. A. *Radiation of Channeling Particles* (in Russian) (Energoatomizdat, Moscow, 1986).
2. Kumakhov, M. A. *Nucl Instrum. Methods* **B48**, 283–286 (1990).
3. Arkadiev, V. A. et al. *Usp. Fiz. Nauk* **157**, 529–539 (1989); (Engl. transl.) *Sov. Phys. Usp.* **32**, 271–276 (1989).
4. Kumakhov, M. A. & Komarov, F. F. *Phys. Rep.* **191**, 289–350 (1990).
5. Kumakhov, M. A., Panin, Yu. N. & Sharov, V. A. *Pisma Zh. tekh. Fiz.* **16**, 24–27 (1990); (Engl. transl.) *Sov. tech. Phys. Lett.* **16**, 371–372 (1990).
6. Kumakhov, M. A., Sharov, V. A. & Kvardakov, V. V. *I.V. Kurchatov Institute of Atomic Energy Preprint* IAE-5412/14 (Moscow, 1991).
7. Kashukuev, N. T. & Chikov, N. F. *Proc. Bulg. Acad. Sci.* **35**, 451–453 (1982).
8. Chen, H. et al. *Nature* **357**, 391–393 (1992).

ACKNOWLEDGEMENTS. We thank V. V. Kuardakov, K. M. Podurets, and D. M. Sokolski for their help during work at the neutron monochromator. We thank V. A. Somerkov and S. S. Shishtein for discussions. Fibres were constructed at the Institute for Glass Technology, Saratov, under the direction of the Institute for Roentgen Optical Systems, Moscow.

Guiding and focusing neutron beams using capillary optics

H. Chen*, R. G. Downing*, D. F. R. Mildner*,
W. M. Gibson†, M. A. Kumakhov‡, I. Yu. Ponomarev‡
& M. V. Gubarev‡

* Chemical Science & Technology Laboratory,
National Institute of Standards and Technology, Gaithersburg,
Maryland 20899, USA

† Center for X-ray Optics, Department of Physics,
State University of New York at Albany, Albany, New York 12222, USA

‡ Institute for Roentgen Optical Systems, World Laboratory and
I. V. Kurchatov Institute of Atomic Energy, Moscow 123182, Russia

A METHOD for focusing X-rays by means of internal reflection inside hollow glass capillaries^{1–3} has also been shown to be capable of guiding a neutron beam⁴, and now a neutron lens using this principle has been demonstrated^{5,6}. Here we study the properties of multiple capillary fibre bundles by measuring the transmission efficiency as a function of bending radius. The fibres used are shown to bend a neutron beam through 20° in a distance of 130 mm, with a transmission of 50%. A simple device made of a few fibres is used to demonstrate the focusing of cold neutrons (wavelength 0.2–0.9 nm) down to a 1-mm^2 spot size. These investigations will form the basis for the design of a neutron lens for use in materials research, or a beam bender capable of dividing and directing a neutron beam towards several experimental stations.

Cold neutrons are now widely used in analysis and characterization of materials⁷, and there is a demand for a neutron probe for quantitative analysis of smaller volumes^{8,9}. A limiting feature of existing probes is the low neutron flux, therefore focusing neutrons using capillary optics³ is desirable. Our method is based on total external reflection from a mirror surface at small grazing angles. Neutrons experiencing multiple reflections inside the highly reflective capillary channels can emerge with little loss in intensity.

The transmission properties of capillaries depend critically on the bending radius, glass composition, surface smoothness and internal diameter. These properties must be understood if

the capillaries are to be used efficiently to construct a high-gain focusing device. For example, quantitative determination of the reflectivity defines a tolerance range for the length of the device; the transmission dependence on the curvature gives an upper limit to the focusing efficiency and the angular range of the device.

The polycapillary fibres used here are made of lead-silica glass with a diameter of 0.4 mm and a hexagonal cross-section⁶. Each fibre contains more than 1,000 capillaries of hexagonal cross-section and has an open area of $\sim 36\%$. Each individual channel has an internal diameter of $6 \mu\text{m}$, an order of magnitude smaller than human hair. An electron beam micrograph of the cross-section of a fibre is shown in Fig. 1. Calculations show that neutrons travelling through capillaries of these dimensions for a length of 250 mm are reflected 100 times on average.

In the bending experiment we held a row of 30 adjacent fibres of length 150 mm stationary at one end, and moved the other

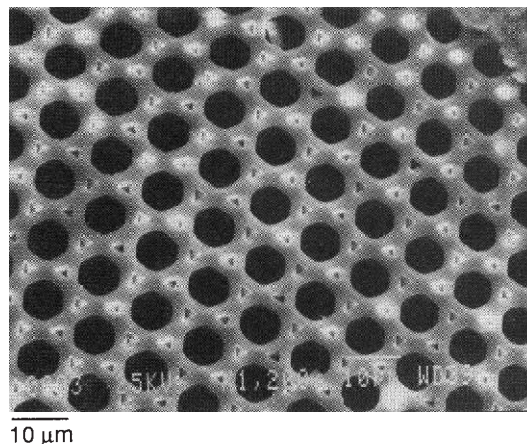


FIG. 1 Electron micrograph of cross-section of a glass polycapillary fibre (courtesy of D. E. Newbury). A fibre contains over 1,000 channels, and each channel has a $6\text{-}\mu\text{m}$ diameter. The open area of the fibre is $\sim 36\%$ of its cross-section.

end by increments to create the curvature (Fig. 2a). The images of the transmitted neutron beams were recorded with a position-sensitive camera (R. G. Downing, manuscript in preparation) at a series of bending radii (Fig. 2b). As the bending radius decreased, the transport efficiency decreased, but even at a bending radius of 0.38 m, the transmission was still remarkably high (~50%). The neutron beam had in this case been deflected by 18.5° over a distance of 130 mm. Some fibres seemed to have better transmission than others; this may result from misalignment or variation in the fibre quality. The curvature created by translating one end only approximates that of a circle, and therefore a more detailed study was required.

In the next experiment, a single fibre was bent to form a segment of a circle with radius R . The transmitted intensity was recorded with the camera as a function of R , and the integrated intensity obtained using image processing software was normalized to the transmitted intensity from a straight fibre. Data collected from two fibres, one 200 mm long with $R = 3.1\text{--}0.15$ m, and the other 167 mm long with $R = 12.6\text{--}0.13$ m, are plotted in Fig. 3. The last data point (radius 0.13 m) from the 167-mm fibre corresponds to the beam being bent by 53.5° with a transmission of ~10%, demonstrating the ability to deflect neutron beams by a large angle over a small distance. Because the beam is polychromatic (wavelength $\lambda = 0.2\text{--}0.9$ nm), the measured intensity is that averaged over the source wavelength spectrum. The theoretical curves were calculated from Kumakhov's capture coefficient³ for a cylindrical guide, adjusted by the loss due to the non-unity reflectivity (0.99 and 0.98 are shown), assuming a square cross-section¹⁰ and allowing only garland

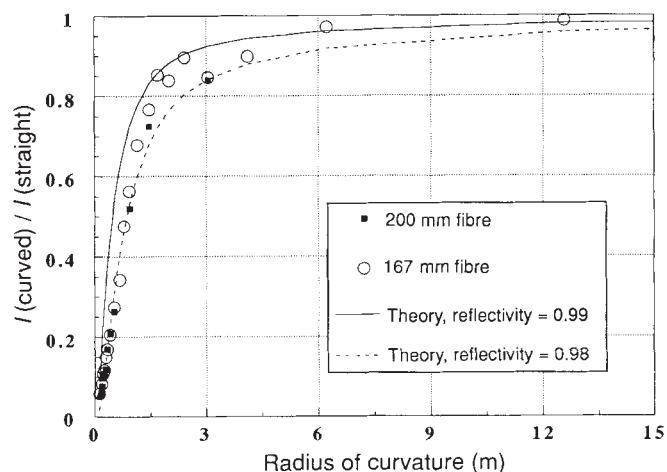


FIG. 3 Measurement of the transmission of neutrons through a polycapillary fibre as a function of the radius of curvature R . Data obtained from fibres 167 mm long and 200 mm long are shown. The transmission is estimated at each wavelength using the Kumakhov model³ for a cylindrical guide, with the parameter $\gamma = R\theta_c^2/2D$ (where $D = 6\text{ }\mu\text{m}$ is the diameter of the individual capillary, and $\theta_c = 9.2\lambda$ is the critical angle for the glass in milliradians per nanometre), multiplied by the factor r^n (where r is the reflectivity of the surface and n is the average number of reflections). We use $r = 0.99$ and 0.98 for the reflection coefficient, and assume that only garland reflections are allowed in a guide having a square cross-section with the same area as the capillary. The calculated transmission is averaged over the incident wavelength spectrum and normalized to that of a straight fibre for comparison with the data. The calculations performed for the 200-mm fibre (not shown) and for the 167-mm fibre indicate that their transmission should be similar, especially at small R . This is verified by two sets of data which agree with each other at small R . The Kumakhov theory is based on a cylindrical guide, and the number of reflections is estimated for a square guide; these approximations may contribute to the deviations of the calculations from the experimental data.

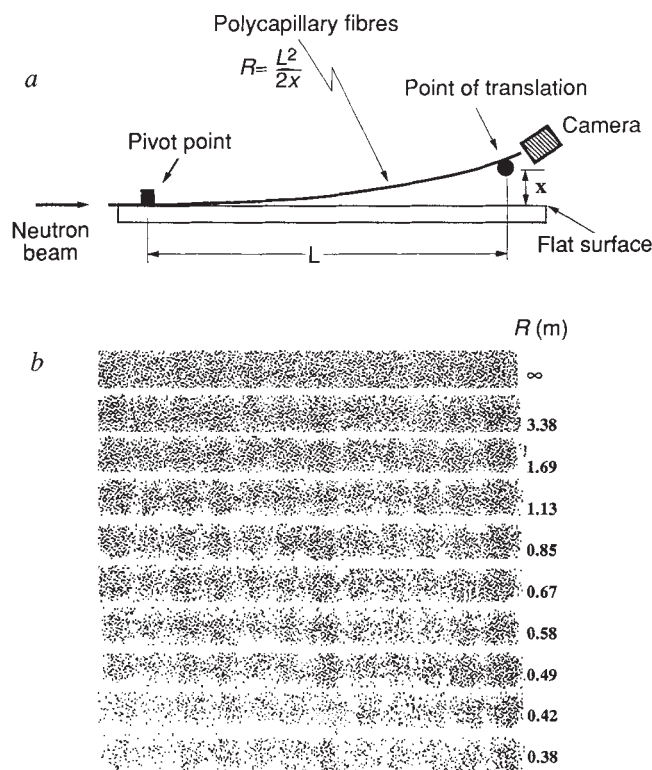


FIG. 2 a, Experimental arrangement for bending a row of polycapillary fibres which are fixed at one end. The other end (at a distance L) is translated by a distance x , such that the curve approximates that of a circle of radius $R = L^2/2x$. b, Neutron beams transmitted through a row of 30 multichannel polycapillary fibres for various bending radii R (only 12 of the adjacent fibre outputs are shown). The transmitted intensity decreases as R decreases. The non-uniformity in transmission may be caused by slight misalignment of the fibres or by differences in the fibre quality. The images were measured by a new position-sensitive detector with a spatial resolution better than $25\text{ }\mu\text{m}$.

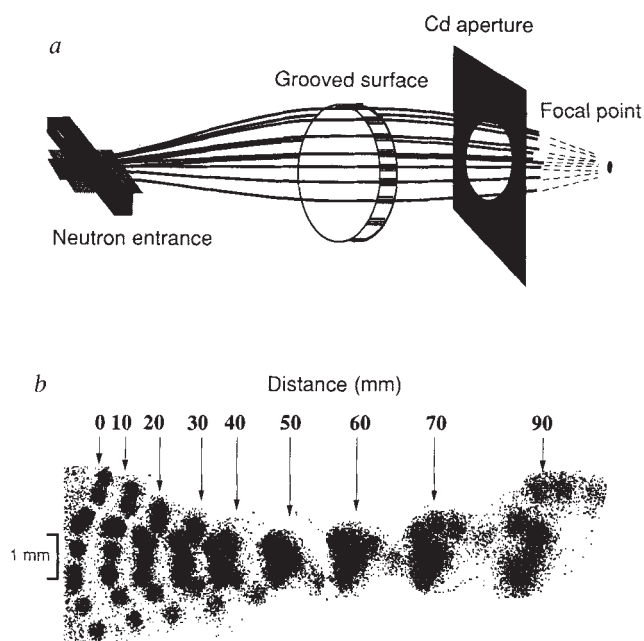


FIG. 4 a, Schematic diagram showing the experimental arrangement used to demonstrate focusing of cold neutrons. b, Images taken along the neutron beam path with the position-sensitive neutron detector at 10 mm intervals from the exit of the polycapillary fibres. Neutrons are focused by an array of nine polycapillary fibres of diameter 0.4 mm , all of which point to the focus at about 50 mm from the exit of the fibres. The focal area is $\sim 1\text{ mm}^2$.

(circumferential) reflections. The theoretical curves resemble the data in form, although deviations are expected because of the strong dependence of the transmission on R , especially at small R . Nevertheless, we conclude that the reflectivity of the capillary surface is close to 0.99 for the range of neutron wavelengths used.

Finally, we used a simple assembly of nine fibres to demonstrate focusing of a polychromatic neutron beam (wavelength 0.2–0.9 nm) of 15 mrad divergence. These fibres were placed side by side on a flat surface. One end of the fibres was fixed for alignment, and the other end spread out onto a cylindrical surface with equally spaced grooves. Each fibre lay along a groove, and the ends of the fibres were passed through a circular cadmium aperture to create conical paths pointing towards a common focal point (Fig. 4a). The position of the aperture was adjusted to determine the location of the focal point. A series of images of the emerging neutron beams recorded at 10-mm intervals from the end of the fibres is shown in Fig. 4b to demonstrate the focusing effect. At 50 mm from the exit, the superimposed neutron beams formed the smallest spot, $\sim 1 \text{ mm}^2$. In an actual application, many hundreds of fibres would be required to achieve a gain of the order of 10 or greater⁶. Careful calculations and designs must be used to optimize the trajectory of each fibre in such an assembly.

Neutron focusing will give neutron absorption techniques greater spatial resolution, allowing them to complement other

techniques for non-destructive analysis of trace elements in materials⁸. It will be straightforward to raster a large specimen, such as a weld or solder joint, across the focused neutron beam, and obtain spatially resolved elemental composition at selected areas. This would allow two-dimensional prompt radiation analysis with submillimetre resolution of elements such as H, B, Li, N and Na. The technique should also lead to three-dimensional imaging for use in mapping elements in microelectronic devices and grain boundaries in materials. Applications to neutron-scattering instrumentation, such as small-angle scattering, may become possible. $\square$

Received 9 January; accepted 8 April 1992.

1. Arkadiev, V. A. *et al. Usp. Fiz. Nauk* **157**, 529–537 (1989); (Engl. transl.) *Sov. Phys. Usp.* **32**, 271–276 (1989).
2. Kumakhov, M. A. *Nucl. Instrum. Methods* **B48**, 283–286 (1990).
3. Kumakhov, M. A. & Komarov, F. F. *Phys. Rep.* **191**, 289–350 (1990).
4. Kumakhov, M. A., Panin, Yu. N. & Sharov, V. A. *Pis'ma Zh. tekh. Fiz.* **16**, 24–27 (1990); (Engl. transl.) *Sov. tech. Phys. Lett.* **16**, 371–372 (1990).
5. Kumakhov, M. A., Sharov, V. A. & Kvardakov, V. V. *I. V. Kurchatov Inst. Atomic Energy IAE-5412/14* (Moscow, 1991).
6. Kumakhov, M. A. & Sharov, V. A. *Nature* **357**, 390–391 (1992).
7. Lindstrom, R. M., Zelsler, R. & Rossbach, M. *J. Radioanalysis Nucl. Chem.* **112**, 321–330 (1987).
8. Lindstrom, R. M., Fleming, R. F. & Rook, H. L. *Microbeam Analysis 1988* (ed. Newbury, D. E. 407–408 (San Francisco Press, 1988).
9. Rossbach, M. *et al. Nucl. Instrum. Methods* **B35**, 181–190 (1988).
10. Mildner, D. F. R. *Nucl. Instrum. Methods* **A292**, 693–699 (1990).

Improved second-harmonic generation from Langmuir–Blodgett films of hemicyanine dyes

G. J. Ashwell*, R. C. Hargreaves*, C. E. Baldwin*,
G. S. Bahra† & C. R. Brown†

* Centre for Molecular Electronics, The Advanced Materials Group, Cranfield Institute of Technology, Cranfield MK43 0AL, UK

† Defence Research Agency, Military Division, RARDE, Fort Halstead, Sevenoaks, Kent TN14 7BP, UK

THE phenomenon of second-harmonic generation (SHG), whereby a material under illumination generates light at twice the incident frequency, is finding increasing use in optical signal processing. The principal requirement for SHG is a non-centrosymmetric structure; in organic materials, this can be achieved through the use of Langmuir–Blodgett films, which offer control of structure at the molecular level. SHG has been demonstrated in films composed of hemicyanine dyes of the general formula $D-C_6H_4-CH=CH-C_5H_4N^+-RX^-$, where D is an electron donor, R is usually a hydrophobic alkyl chain and X^- is a counterion such as Br^- or I^- . The frequency-doubling properties of these films are sensitive to the choice of donor group^{1,2}, the extent of molecular aggregation^{3,4} and the type of packing^{5–7}. Mixed films, in which the dye molecules are interspersed with an inert phase, have been used to reduce aggregation and consequently enhance SHG^{3,8}, but these films have the potential disadvantage of phase separation. Here we show that the use of an amphiphilic anion as both counterion and spacer molecule gives rise to an ordered segregation of the hemicyanine chromophores, and greatly enhanced SHG.

The dye was obtained by the metathesis of E-N-octadecyl-4-(2-(4-dimethylaminophenyl)ethenyl)pyridinium iodide and sodium octadecylsulphate, spread in a 1:1 ratio from a dilute methanol–chloroform solution onto the pure water subphase of a NIMA Technology Langmuir–Blodgett trough. We assumed that the water-soluble ions, Na^+ and I^- , dissolve into the subphase, and this is supported by analysis.

The floating monolayer was compressed to $30\text{--}40 \text{ mN m}^{-1}$ and transferred to a hydrophilically treated glass slide at a rate of 5 mm min^{-1} . The isotherm is steeper than that of the corresponding iodide (Fig. 1) and exhibits an alignment anomaly at about 18 mN m^{-1} . The area per molecule at deposition agrees well with the combined van der Waals cross-section of the cation and anion hydrophobic chains (0.42 nm^2), indicating that the legs are upright and closely packed (see Fig. 2). Note that the dye is very slightly water soluble and thus the area per molecule may be larger than the values calculated from the volume and concentration of solution spread on the subphase.

SHG was measured in transmission with a Y-cut quartz plate as reference and with the Nd:YAG laser beam at an angle (θ) of 45° to the film⁵. The SH intensity ($I_{2\omega}$) from the monolayer is given by the equation:

$$I_{2\omega} \propto \frac{(\chi_{\theta}^{(2)} I_{\omega})^2}{n_{\omega}^2 n_{2\omega}} \quad (1)$$

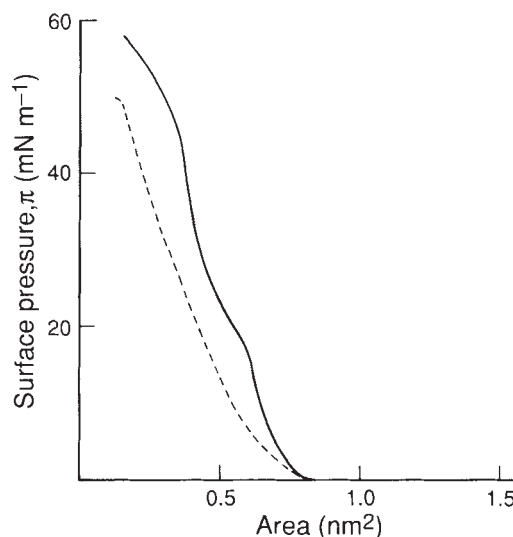


FIG. 1 Surface pressure (π) against molecular area isotherms of E-N-octadecyl-4-(2-(4-dimethylaminophenyl)ethenyl)pyridinium octadecylsulphate (solid line) and iodide (broken line).

where n_ω and $n_{2\omega}$ are the refractive indices at 1,064 nm and 532 nm, respectively, l is the film thickness, and $\chi_\theta^{(2)}$ is the measured susceptibility, uncorrected for angular dependence. The SHG was investigated as a function of the incident polarization (see Fig. 3) and the data analysed using equations (2) to (4) below, where N is the number of molecules per unit volume and $f_\omega = (n_\omega^2 + 2)/3$, a local field correction factor. We assumed that the chromophores have a common tilt angle, ϕ (Fig. 2), with a random azimuthal distribution and that the second-order molecular hyperpolarizability (β) is dominated by the component along the intramolecular donor- π -acceptor axis.

$$\frac{I_{2\omega}(p \rightarrow p)}{I_{2\omega}(s \rightarrow p)} = \frac{(\chi_{zzz}^{(2)} \sin^3 \theta + 3\chi_{zxx}^{(2)} \sin \theta \cos^2 \theta)^2}{(\chi_{zxx}^{(2)} \sin \theta)^2} \quad (2)$$

$$\chi_{zzz}^{(2)} = N f_{2\omega} (f_\omega)^2 \beta \cos^3 \phi \quad (3)$$

$$\chi_{zxx}^{(2)} = \frac{1}{2} N f_{2\omega} (f_\omega)^2 \beta \cos \phi \sin^2 \phi \quad (4)$$

The polarization-dependent second-harmonic intensities are in the order $I_{2\omega}(p \rightarrow p) \gg I_{2\omega}(s \rightarrow p)$, with $I_{2\omega}(p \rightarrow s) \sim I_{2\omega}(s \rightarrow s) \sim 0$; the ratio of intensities from Fig. 3 provide a tilt angle of 18° and using this, a film thickness of 3.4 nm has been obtained by molecular modelling. The refractive indices are unknown, but from surface plasmon resonance studies on the pyridinium zwitterion, Z- β -(1-octadecyl-4-pyridinium)- α -cyano-4-styryl-dicyanomethanide, whose structure is similar, n is taken to be ~ 1.7 . Thus, by comparison with the signal from the quartz reference (d_{11} , 0.5 pm V^{-1}) and using the values above, we obtain $\chi_{zxx}^{(2)} = 30\text{--}80 \text{ pm V}^{-1}$, $\chi_{zzz}^{(2)} = 500\text{--}1,500 \text{ pm V}^{-1}$ and $\beta = 2\text{--}5 \times 10^{-37} \text{ m}^4 \text{ V}^{-1}$.

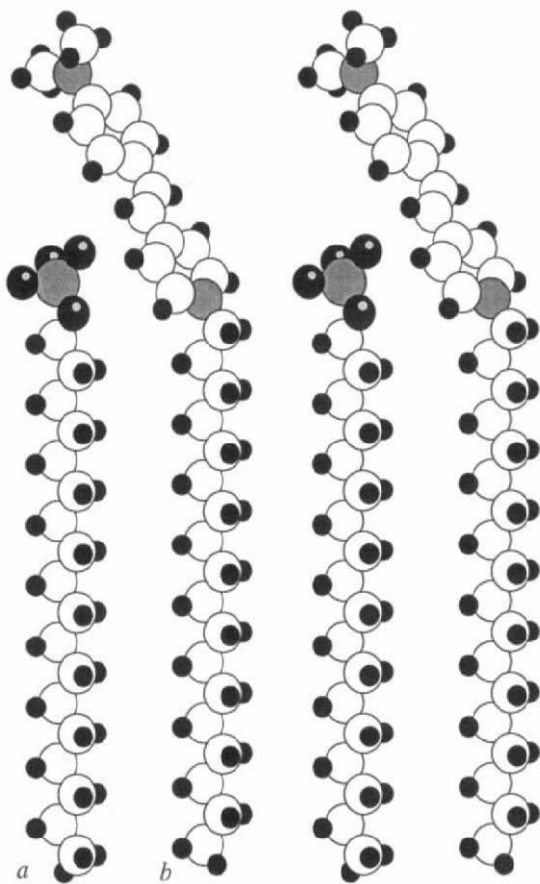


FIG. 2 Possible packing arrangement of E-N-octadecyl-4-(2-(4-dimethylaminophenyl)ethenyl)pyridinium octadecylsulphate within the Langmuir-Blodgett film. a, Anion; b, cation.

TABLE 1 Anion dependence of the second-order susceptibility and tilt angle (from SHG data) and absorbance maximum of Langmuir-Blodgett films of a hemicyanine chromophore

Hemicyanine dye	$\chi_{zzz}^{(2)}$ (pm V^{-1})	ϕ ($^\circ$)	λ_{max} (nm)
E-N-octadecyl-4-(2-(4-dimethylaminophenyl)ethenyl)pyridinium octadecylsulphate	500–1,500	18	460
iodide	90–150	41	460
bromide	70–90	40	460

To convert the units of $\chi^{(2)}$ to e.s.u. divide by $4\pi/(3 \times 10^4)$.

The SHG from films of the hemicyanine octadecylsulphate is compared in Fig. 4 with signals from representative examples of optically nonlinear materials studied at Cranfield. We note, for example, that the weakest intensity from films of the octadecylsulphate salt is equivalent to the signal from ~ 30 Z-type layers of the pyridinium zwitterion and five Z-type layers of the quinolinium congener (Fig. 4). We also note that the second-harmonic intensity of the hemicyanine dye is strongly anion-dependent (Table 1), the second-order susceptibility $\chi_{zzz}^{(2)}$ of the bromide and iodide being 70–90 and 90–150 pm V^{-1} , respectively.

Use of the amphiphilic anion was prompted by the variable behaviour of the related hemicyanine, E-N-docosyl-4-(2-(4-dimethylaminophenyl)ethenyl)pyridinium bromide, first reported by Girling *et al.*^{8–10} and extensively studied elsewhere^{11–20}; its properties are dependent on the extent of molecular aggregation which can be suppressed by diluting the chromophore layer^{3,8}. Notably, for a 1:4 film of the hemicyanine and behenic acid, Schildkraut *et al.*³ have disclosed a bathochromic shift of the absorption band from 339 nm to 477 nm and a 44-fold increase in the second-harmonic intensity, the improvement being due in part to resonant enhancement.

Mixed films have the potential disadvantage of phase separation, so an alternative method of chromophore segregation has been explored by using a spacer that constitutes part of the molecule. For the hemicyanine octadecylsulphate (see Fig. 4 for its molecular structure), the fatty alkyl chains of the cationic dye and amphiphilic counterion should pack side-by-side, with the negatively charged $-\text{OSO}_3^-$ groups interleaving the positively charged pyridinium rings. Improved SHG has been realized by this method (Table 1), but Langmuir-Blodgett films of

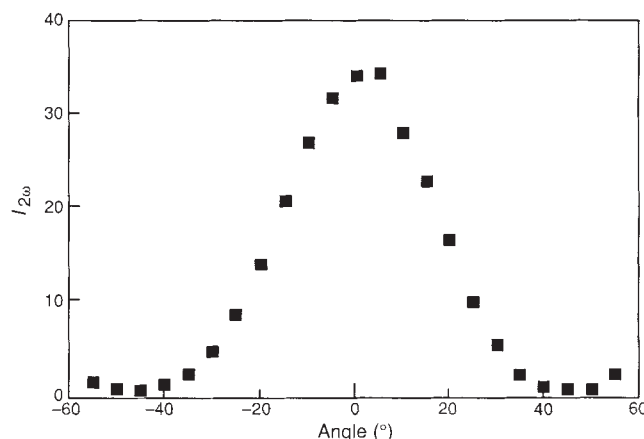


FIG. 3 Polarization dependence of the SHG (arbitrary units) from a monolayer film of E-N-octadecyl-4-(2-(4-dimethylaminophenyl)ethenyl)pyridinium octadecylsulphate, the peak and troughs corresponding to the signal from p -polarized and s -polarized fundamental beams, respectively. The abscissa corresponds to the angle of rotation of the half-wave plate.

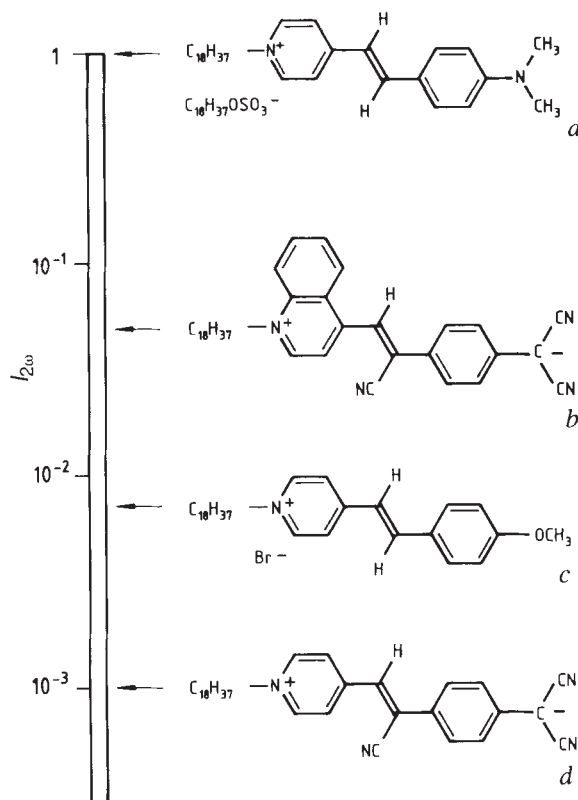


FIG. 4 Comparison of the minimum second-harmonic intensity (arbitrary units) from monolayer films of E-N-octadecyl-4-(2-(4-dimethylamino)phenyl)ethenylpyridinium octadecylsulphate (a) with typical signals from films of related optically nonlinear materials mentioned in the text: Z-β-(1-octadecyl-4-quinolinium)-α-cyano-4-styryldicyanomethanide (b); E-N-octadecyl-4-(2-(4-methoxyphenyl)ethenyl)pyridinium bromide (c); Z-β-(1-octadecyl-4-pyridinium)-α-cyano-4-styryldicyanomethanide (d).

the bromide, iodide and octadecylsulphate have almost identical visible spectra with maxima at 460 nm and peak absorbances of 0.008 per layer (compare with $\lambda_{\max} \approx 477$ nm for the non-aggregated phase reported by Schidkraut *et al.*³). Similarly, from films of the methoxy congener, E-N-octadecyl-4-(2-(4-methoxyphenyl)ethenyl)pyridinium bromide, enhanced SHG has been realized by replacing the counterion with octadecylsulphate, again without change to the visible spectrum. Thus, unlike the behaviour of the mixed hemicyanine/behenic acid films³, the stronger signals are not attributable to resonant enhancement as the spectral overlap at the harmonic wavelength is the same in each case.

The improved signal is probably due to chromophore ordering by way of the alternating amphiphilic cations and amphiphilic anions, the dominant interaction in the aligned film being between the oppositely charged head groups with a weaker van der Waals interaction between the fatty hydrophobic chains. But defects may arise from an antiparallel head-to-tail overlap of the donor and acceptor parts of adjacent chromophores, the SHG being cancelled by the centrosymmetric alignment in such regions of the film. This seems to be particularly relevant to the pyridinium zwitterion (Fig. 4) whose signal is only 2% of that of the bulkier quinolinium congener, yet from a theoretical basis the hyperpolarizabilities should be similar. It is also probably the reason for the reduced SHG from films of the hemicyanine halides, compared with the octadecylsulphate, as the likelihood of antiparallel alignment is greater when the donor- π -acceptor chromophores are not separated. □

Received 16 January; accepted 9 April 1992.

1. Lupo, D. *et al.* *J. opt. Soc. Am.* **B5**, 300-308 (1988).
2. Bubeck, C. *et al.* *Angew. Chem. Int. Edn Engl.* **3**, 54-58 (1991).
3. Schidkraut, J. S., Penner, T. L., Willand, C. S. & Ullman, A. *Opt. Lett.* **13**, 134-136 (1988).
4. Steinhoff, R., Chi, L. F., Marowsky, G. & Möbius, D. *J. opt. Soc. Am.* **B6**, 843-847 (1989).
5. Ashwell, G. J., Dawney, E. J. C., Kuczyński, A. P. & Martin, P. J. *SPIE Int. Soc. opt. Engng* **1361**, 589-598 (1991).
6. Ashwell, G. J. *et al.* *Mat. Res. Soc. Symp. Proc.* **247**, 787-792 (1992).
7. Anderson, B. L. *et al.* *Thin Solid Films* **179**, 413-421 (1989).
8. Girling, I. R. *et al.* *J. opt. Soc. Am.* **B4**, 950-955 (1987).
9. Girling, I. R. *et al.* *Thin Solid Films* **132**, 101-112 (1985).
10. Cross, G. H. *et al.* *Thin Solid Films* **156**, 39-52 (1988).
11. Young, M. C. *et al.* *Thin Solid Films* **182**, 319-332 (1989).
12. Kowal, S. T., Hayden, L. M. & Selfridge, R. H. *SPIE Int. Soc. opt. Engng* **682**, 103-108 (1986).
13. Hayden, L. M. *et al.* *Thin Solid Films* **160**, 379-388 (1988).
14. Selfridge, R. H. *et al.* *Thin Solid Films* **160**, 471-476 (1988).
15. Stroeve, P., Saperstein, D. D. & Rabolt, J. F. *Thin Solid Films* **179**, 529-534 (1989).
16. Stroeve, P., Saperstein, D. D. & Rabolt, J. F. *J. chem. Phys.* **92**, 6958-6967 (1990).
17. Ancelin, H. *et al.* *Langmuir* **6**, 172-177 (1990).
18. Shirota, K., Kajikawa, K., Takezoe, H. & Fukuda, A. *Jap. J. appl. Phys.* **29**, 750-755 (1990).
19. Kajikawa, K., Shirota, K., Takezoe, H. & Fukuda, A. *Jap. J. appl. Phys.* **29**, 913-917 (1990).
20. Kajikawa, K., Kigata, K., Takezoe, H. & Fukuda, A. *Molec. Cryst. Liq. Cryst.* **A182**, 91-101 (1990).

ACKNOWLEDGEMENTS. G.J.A. thanks the SERC Molecular Electronics Committee for support of the nonlinear optics programme and the Defence Research Agency for a studentship to R.C.H.

An electrochemical route to pre-shaped ceramic bodies

P. J. Mitchell* & G. D. Wilcox†

* Department of Chemistry and † Institute of Polymer Technology and Materials Engineering, Loughborough University of Technology, Loughborough LE11 3TU, UK

A SIMPLE technique for producing complex ceramic structures would be valuable for technological applications involving these materials. We have developed a method for fabricating 'preshaped' ceramic bodies using an electrochemical cell to deposit a ceramic precursor material onto the cathode. The deposit may be readily detached from the electrode, while retaining its shape, before the final calcination process. Alternatively, it may be preferable to retain the material on the cathode throughout the subsequent processing steps. Whereas previous attempts to generate ceramics electrochemically resulted in the formation of powders^{1,2} or thin adherent films³⁻⁶, our technique offers a high degree of morphological control, can produce porous or compact structures, and can be readily extended to multilayer and composite materials. Definition of more complex structures becomes possible through shaping and masking of the electrodes. The application of this inherently simple technique to functional ceramics (for example, piezoelectrics, semiconductors and superconductors) should prove a promising area for future investigation.

The electrolytic decomposition of water results in the local generation of alkaline conditions at the cathode. This is attributable to the reactions $2\text{H}^+ + 2\text{e}^- \rightarrow \text{H}_2$ in acid solution, or $2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{H}_2 + 2\text{OH}^-$ in neutral or basic media. In aerated solutions it is likely that there will be a contribution from the reduction of oxygen.

Alkaline conditions will promote the precipitation of metallic hydroxides from a suitable electrolyte solution. If the local alkaline environment is not disrupted by convective or other forces then the accumulated solid phase remains near to the cathode and forms a gel-like structure. To promote such conditions, it is essential that the gaseous hydrogen evolution from the above process is minimized and preferably eliminated. This may be achieved by use of a hydrogen-sorbing cathode material such as palladium, iron-titanium alloys or rare-earth-nickel alloys. Uniformity of deposit depends primarily on the current distribution, which is in turn determined by the electric field. As the deposited material is not electronically conducting, it introduces a growing resistance to the flow of current as its thickness increases. This tends to adjust the electric field so as to force base generation to occur at uncovered sites, thereby promoting an even deposit thickness.

The physical characteristics of the deposit depend on the concentration of the metallic species in the electrolyte and on the current density at the cathode. This is presumably because there is a zone of precipitation whose rate of advancement is determined by the balance between the diffusion fluxes of the incoming metallic cation and the outgoing hydroxyl anion. Flux is described by the one-dimensional form of Fick's first law thus:

$$\text{Flux} = -D_i \, dC_i/dx$$

where D_i is the diffusion coefficient and C_i the concentration of species i (M^+ or OH^-). The metal ion flux will respond to changes in bulk concentration, whereas the flux of hydroxyl ion will be dependent on its concentration at the electrode-electrolyte interface. This value can be calculated from a combination of Faraday's law and Fick's law

$$C_{OH} = (2j/F)\sqrt{t/(\pi D_{OH})}$$

where F is the Faraday constant, j the cathode current density and t the time elapsed after imposition of the current.

Thickness of deposit is governed by the ease of transport of reactant species through the precipitating material. As the density of the deposit increases, it is likely that field-driven ionic migration will supplement diffusion and eventually become the predominant transport mechanism. It is therefore essential that a sufficiently negative potential can be maintained at the cathode surface to ensure a constant current density. This involves increasing the overall cell potential to compensate for the potential drop in the depositing layer. Thicknesses up to 0.3 mm (and probably greater) can be achieved with time. Multilayer and multicomponent systems can be produced by withdrawing the integral cathode and associated deposit from one electrolyte solution, transferring to another and resuming the electrolysis. The growing deposit easily entraps suspended particles of other solid materials, producing a dispersion or composite structure.

A wide range of electrolyte compositions and cathode current densities can be used in an electrochemical cell which will produce deposits of the types described. Useful deposits or structures have been obtained from electrolytes containing non-dischargeable metallic cations (species that cannot be electrochemically reduced to the metal from aqueous electrolytes) including aluminium, calcium, magnesium, titanium and zirconium. Anionic counterparts for the above have included nitrate, chloride and sulphate. Density, porosity and crystal size vary considerably with electrolysis conditions. Higher electrolyte concentrations tend to favour a densely packed deposit which adheres strongly to the electrode. Lower concentrations generally result in the gelatinous product which, depending on rate

of dehydration, may or may not be porous and is often detachable in the 'wet' form. Binary and ternary compositions are produced by suitable mixtures of electrolyte. Dischargeable metals can be co-deposited with non-dischargeable metals: examples of such compositions are copper, cobalt or nickel with magnesium, and lead with calcium. Particulate platinum metal has been introduced into a growing aluminium hydroxide precipitate from a surfactant-stabilized platinum sol containing aluminium sulphate. A dispersion of small submicrometre metal particulates and larger (10–50 μm) aggregates in alumina was obtained. We have not yet investigated the effect of temperature; all results so far have been obtained at 25 °C.

Our results establish the feasibility of producing thick, pre-shaped deposits of ceramic precursor materials on a hydrogen-sorbing substrate (in this case palladium). Figure 1 shows a tube fabricated in two stages with the inner and outer layers composed of aluminium and magnesium oxides, respectively. The basic procedure was as follows. A palladium wire of diameter 0.4 mm and length 10 mm was used as cathode, and immersed in a glass cell containing ~10 ml of 0.5 mol aluminium sulphate. A cylindrical anode of platinum gauze was used, and all potentials were measured against a mercury/mercurous sulphate reference electrode. Galvanostatic control was achieved by simple modification of a commercially available potentiostat (Thompson Electrochem Ministat). The temperature was maintained at 25 °C. The deposition was done at 20 mA cm^{-2} for 150 s. The circuit was then broken, and the cathode was withdrawn and placed in a second cell containing 0.5 mol magnesium sulphate. Electrolysis was resumed, at the same current density, for a further 150 s. The deposit was then withdrawn from the cell and simply eased from the cathode with tweezers. The resulting tube was a self-supporting, translucent, fine-grained structure composed of two distinct layers. A final calcining process was carried out at 250 °C, in air, for 1 h. This resulted in the open porous body of Fig. 1. A more compact crystalline form can be produced if the calcining temperature is increased to above 400 °C.

For more complex structures, we recognize that shaping or masking of the electrode will be necessary to define the final form of the resulting body. This should not be a particular problem, as it is a standard procedure in the electroplating and electroforming industry. Masking uses chemically inert and electrically resistive compounds which can be applied particularly accurately by photoresist methods. We envisage that once a design has been specified, a master 'former' or 'mandril' would be produced. This would allow the production of many thousands of pre-shaped ceramic bodies for minimal initial capital outlay.

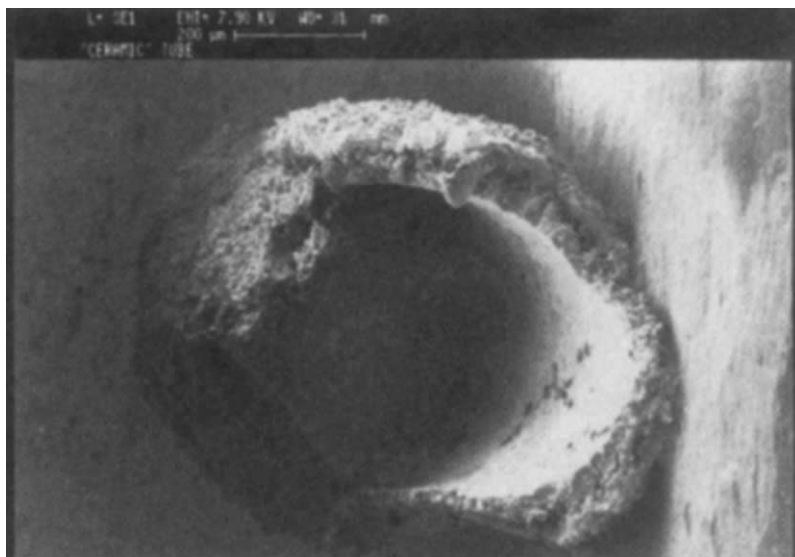


FIG. 1 Two-layer ceramic tube composed of aluminium oxide (inner) and magnesium oxide (outer).

Experiments with more common electrode materials, including steel and copper, have shown that at sufficiently low current densities thin, uniform, adherent deposits can be generated at their surfaces. We have noticed that these coatings alter both the physical and the chemical surface properties of the substrate, and we are investigating this effect.

It is now important to select specific applications and materials for further research. Many possibilities exist in the fields of engineering (in wear resistance, heat resistance and cutting tools, for example) and functional ceramics (such as piezoelectric material, semiconductors, superconductors, ion conductors and ferrites). □

Received 25 November 1991; accepted 3 April 1992.

- Switzer, J. A. *Am. Ceram. Soc. Bull.* **66**, 1521–1524 (1987).
- Coyle, R. T. & Switzer, J. A. US Patent 4,882,014 (1989).
- Slezak, P. & Wieckowski, A. *J. electrochem. Soc.* **138**, 1038 (1991).
- Bhattacharya, R. N., Noufi, R., Roybal, L. L. & Ahrenkil, R. K. *J. electrochem. Soc.* **138**, 1643–1645 (1991).
- Switzer, J. A., Shane, M. J. & Phillips, R. J. *Science* **247**, 444–445 (1990).
- Lashmore, D. S. & Dariel, M. P. *J. electrochem. Soc.* **135**, 1218–1221 (1988).

Episodic atmospheric nitrogen deposition to oligotrophic oceans

N. J. P. Owens*, J. N. Galloway† & R. A. Duce‡

*Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK

†Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia 22903, USA

‡College of Geosciences and Maritime Studies, Texas A&M University, College Station, Texas 77843, USA

PHYTOPLANKTON production is generally thought to be limited by nitrogen availability over large regions of the ocean, particularly oligotrophic areas. The largest source of nitrogen to these regions is transport from deep waters. Atmospheric deposition is also a source of nitrogen to the oceans^{1,2}, but previous assessments of these inputs, based on calculations of annual deposition^{3,4}, suggest that they are of minor importance for oceanic production. Using a nine-year record of nitrogen deposition, we re-evaluate here the contribution of atmospheric nitrogen inputs to the ocean and show that these can contribute to an important proportion of 'new production' during episodic events. Furthermore, as human activities such as fossil-fuel burning contribute significant quantities of inorganic nitrogen to the atmosphere^{5,6}, we suggest that the atmospheric input of nitrogen to the oceans is a route whereby these activities might directly influence (by increasing) oceanic productivity.

Data on the wet deposition of nitrate and ammonium are available for Bermuda (North Atlantic)⁷, and for Amsterdam Island (South Indian Ocean)⁸ for a nine-year period. The data were obtained using methods described previously^{7,8}. The record (Table 1) shows that the median and maximum rates of episodic nitrogen deposition at Bermuda are several times larger than at Amsterdam Island. An event may last for periods of hours to days. At Bermuda, the typical event is 1 hour to 2 days. On a seasonal basis, the median deposition per event at Bermuda is essentially the same in the winter and summer (83.4 and 86.4 $\mu\text{mol m}^{-2}$ per event, respectively), whereas at Amsterdam Island, the median deposition is slightly larger in the summer than in the winter (28.0 and 16.9 $\mu\text{mol m}^{-2}$ per event, respectively). Events are as likely in the summer as in the winter at both Amsterdam Island and Bermuda.

The difference in deposition between the two sites is due to their different distances from continental sources of atmospheric nitrogen. Bermuda is close to continental regions surrounding the North Atlantic Ocean, and the atmospheric deposition is

TABLE 1 Summary of inorganic nitrogen deposition

	Bermuda	Amsterdam Island
Position	32° 18' N 64° 45' W	37° 00' S 79° 00' E
Sampling period	1 April 1980 to 14 August 1989	17 May 1980 to 21 July 1989
Number of events sampled	512	249
Median deposition ($\mu\text{mol N m}^{-2}$ per event)	85.1	22.4
Maximum deposition ($\mu\text{mol N m}^{-2}$ per event)	1,470	510
Minimum deposition ($\mu\text{mol N m}^{-2}$ per event)	0.70	0.34

partly anthropogenic (such as nitrogen oxides and ammonia) and partly 'natural', subsequently transported to, and deposited from, the marine atmosphere. Because anthropogenic sources dominate in eastern North America⁹, we assume that the difference between nitrogen deposition at Bermuda and Amsterdam Island is primarily anthropogenic.

The impact of atmospheric deposition of nitrogen on phytoplankton productivity has previously been examined by comparing the annual inputs with the annual phytoplankton requirement^{3,4}. This approach suggests that the atmospheric inputs are considerably less than the requirement and not an important source. It is evident from Fig. 1, however, that substantial inputs of nitrogen occur episodically. We have investigated the possible impact of these events by comparing the nine-year frequency distribution of the episodic events at Bermuda with a range of daily nitrogen requirement (Fig. 2).

We have assumed a daily phytoplankton production for the Sargasso Sea of 100–500 $\text{mg C m}^{-2} \text{ day}^{-1}$. These values are near the top of the range for phytoplankton production of oligotrophic oceans^{10,11} and will lead to a conservatively low

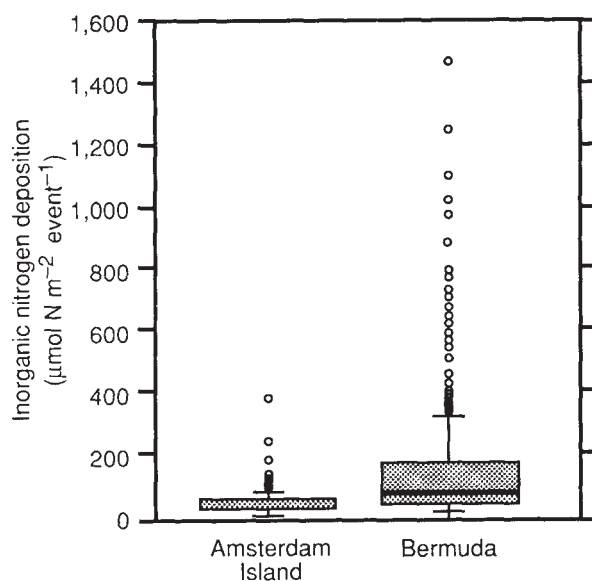


FIG. 1 Percentile distribution of deposition events ($\mu\text{mol N m}^{-2}$) of nitrate and ammonium for the period 1 April 1980 to 14 August 1989 for Bermuda (North Atlantic Ocean) and 17 May 1980 to 21 July 1989 for Amsterdam Island (South Indian Ocean). Collection and measurements were made as described in refs 3 and 4. The limits shown are: the median event (line within box; obscured in the Amsterdam Island data), 25% to 75% of events (shaded box) and 10% to 90% of events (error bars). Events outside these limits are shown as individual points.

estimate of the importance of the atmospheric inputs. Direct measurements of nitrogen assimilation suggest that the proportion of nitrogen derived from external sources ('new nitrogen'¹²) may be up to 0.2 for oligotrophic oceans, although the high degree of recycling typical of such areas is commonly characterized by 'new production' (f ratio) as low as 0.05 (ref. 13). By applying the higher proportion (0.2) to the range of production, and assuming a Redfield ratio (C/N) of 6.6, we calculate a range of 'new nitrogen' assimilation of 216–1,081 $\mu\text{mol N m}^{-2} \text{ day}^{-1}$. The use of the higher f ratio tends to overestimate the 'new nitrogen' assimilation, perhaps by as much as fourfold, thus further underestimating the importance of the atmospheric inputs.

Figure 2 shows that more than 20% of the episodic inputs at Bermuda could individually supply enough nitrogen to meet 100% of the minimum 'new nitrogen' daily demand; roughly 40% of the events could supply at least 50% of the demand. If the maximum 'new nitrogen' requirement is assumed, less than 5% of the events could supply sufficient nitrogen to satisfy the demand. Data on primary production around Bermuda show, however, that the lower level of production assumed is more typical for eight months of the year, with production levels of the order of the maximum occurring only during January to April, and then only periodically¹⁴. Assuming a similar level of primary production around Amsterdam Island, it is apparent that only the most extreme events could supply more than 50% of even the minimum nitrogen demand.

Clearly there is an inconsistency between the units of time used for the nitrogen demand (day^{-1}) and nitrogen inputs (per event). Given that an event typically lasts between a few hours and two days, however, we consider that these timescales are sufficiently close to warrant comparison, especially as the atmospheric inputs are underestimated in this analysis because only wet deposition has been considered. Dry deposition also introduces inorganic nitrogen, with inputs as high as 25% of the wet deposition in remote marine areas⁶. In addition, low-molecular-weight organic nitrogen compounds (such as amines), which can be used by the phytoplankton, are also found in nanomolar concentrations in rain water¹⁵, but have been ignored in this analysis. Knap *et al.*³ have shown that the organic nitrogen

fraction may amount to 50% of the inorganic fraction. If all of this is usable the contribution of atmospheric nitrogen could potentially be double that suggested here, not including the underestimates outlined above.

Direct observations have been made of a transient phytoplankton bloom initiated and supported by an injection of inorganic nitrogen in the Sargasso Sea¹⁶. The bloom comprised predominantly *Synechococcus* and coincided with an increase in dissolved nitrate concentrations to 21 nmol l^{-1} . Although this bloom followed a rain event, it was concluded that the nitrogen increase was not derived from the atmospheric input but rather resulted from an input of nutrient-rich deep water. We have examined our data in the context of these direct observations. For an upper mixed layer of 25 m and no diffusive losses, the median input of nitrogen from the Bermuda data (85.1 $\mu\text{mol m}^{-2}$ per event) would increase the inorganic nitrogen concentrations by only 3.4 nmol l^{-1} if distributed throughout the mixed layer. In contrast, the maximum input (1,470 $\mu\text{mol m}^{-2}$ per event) represents an increase of up to 58.8 nmol l^{-1} . Events of such importance are rare. During the directly observed bloom, the calculated nitrogen demand was 417 $\mu\text{mol m}^{-2}$ over a five-day period. The nine-year record shows that deposition events introduced such quantities of nitrogen on average three times a year. Clearly, smaller input events would only support a lower nitrogen demand, but it is evident that these events occur frequently throughout the year. For example, an average of 51 deposition events (range 24–84, s.d. 19.4) occur at Bermuda each year. Of these, an average of 22 events (range 11–35, s.d. 8.7) are sufficient to supply more than 50% of the calculated daily 'new nitrogen' demand.

Horizontal dispersion, on the other hand, acts to reduce the effect of the inputs. For a point-source input that increased the nitrogen concentration of the surface mixed layer to 21 nmol l^{-1} (equivalent to the directly observed bloom), and horizontal dispersion of $5 \times 10^3 \text{ cm}^2 \text{ s}^{-1}$ for a horizontal scale of 1 km (ref. 17), dispersion would reduce the nitrate concentration by $\sim 0.9 \text{ nmol l}^{-1} \text{ day}^{-1}$. The removal rate by phytoplankton was equivalent to $\sim 3 \text{ nmol l}^{-1} \text{ day}^{-1}$, suggesting that even in the unrealistic case of a point source input, phytoplankton assimilation can overcome horizontal dispersion. The inputs are not point sources; they are distributed over spatial scales from tens to hundreds of kilometres (examples might be individual thunderstorms and major frontal storms, respectively). Over such spatial scales, horizontal dispersion is insignificant.

Glover *et al.*¹⁶ showed that the increase in number of *Synechococcus* cells, initiated by an input of nitrogen, was consistent with stoichiometric utilization of the nitrogen introduced. The stimulation of primary production was rapidly transferred into higher trophic levels, with 70% of the *Synechococcus* cells consumed by grazers within two days. Experiments in which rain water was added to sea water have shown increases in chlorophyll concentrations by more than 100% within two to three days^{2,18}. It is evident, therefore, that both producers and consumers in oligotrophic systems are able to respond rapidly to episodic inputs of nitrogen. We suggest that the frequent episodic deposition events stimulate a series of such 'blooms' throughout the year and conclude that the atmospheric inputs of nitrogen are more significant than hitherto supposed.

Our analysis shows that atmospheric inputs are a frequent, and quantitatively large, episodic source of 'new nitrogen' in nitrogen-depleted oceans. The influence will be most pronounced in areas that are most affected by continental inputs (compare Bermuda and Amsterdam Island; see also ref. 1). Given that extremely low levels of nitrogen, comparable to those in oligotrophic oceans, are a feature of seas only tens of kilometres from land (for example, most of the North Sea in the summer¹⁹) this influence is not confined to distant oceanic regions; indeed, the nearer-shore regions, being closer to the sources, may be subject to greater inputs than those considered here.

These findings have important implications for our under-

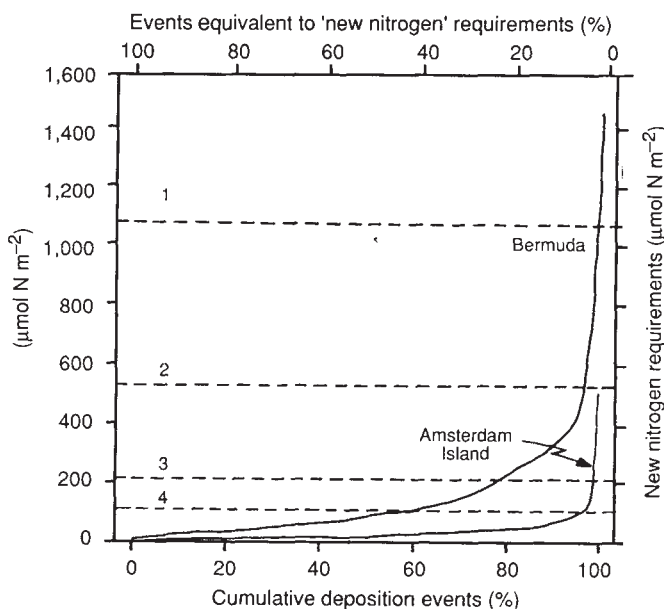


FIG. 2 Cumulative frequency (%) of wet deposition of inorganic nitrogen at Bermuda and Amsterdam Island, and the percentage of events contributing to a range of estimated (see text) 'new nitrogen demand' (horizontal dotted lines 1–4). Lines 1 and 2 are the maximum and 50% of maximum 'new nitrogen' requirement; 3 and 4 are the minimum and 50% minimum.

standing and the measurement of 'new production' in oligotrophic seas. If atmospheric inputs are ignored, as they usually are, we conclude that 'new production' will be underestimated. Platt and Harrison¹⁴ also concluded that short-term enhancement of inorganic nitrogen concentrations, in general, accounts for a disproportionately high fraction of production in nitrogen-limited systems. Furthermore, because

anthropogenic sources contribute to the deposition of atmospheric nitrogen to the ocean, these episodic inputs provide a mechanism whereby human activities may directly influence ocean phytoplankton productivity. Given that anthropogenic NO_x emissions have increased by almost 100% over the past three decades⁵, the possibility exists that phytoplankton production could also have been enhanced. □

Received 9 December 1991; accepted 3 April 1992.

- Fanning, K. A. *Nature* **339**, 460–463 (1989).
- Paerl, H. W. *Nature* **315**, 747–749 (1985).
- Knap, A., Jickells, T., Pszenny, A. & Galloway, J. N. *Nature* **320**, 158–160 (1986).
- Duce, R. A. in *The Role of Air–Sea Exchange in Geochemical Cycling* (ed. Buat-Menard, P.) 497–529 (Reidel, Dordrecht, 1986).
- Hameed, S. & Dignon, J. *Atmos. Envir.* **22**, 441–449 (1988).
- Duce, R. A. *et al.* *Global biogeochem. Cycles* **5**, 193–259 (1991).
- Galloway, J. N., Tokos, J. J., Knap, A. H. & Keene, W. C. *Tellus* **B40**, 178–188 (1988).
- Moody, J. L. *et al.* *J. geophys. Res.* **96**, 20769–20786 (1991).
- Galloway, J. N. & Rodhe, H. *Proc. R. Soc. Edin.* **B97**, 61–80 (1991).
- Koblentz-Mishke, O. J., Volkovinsky, V. V. & Kabanova, J. G. in *Scientific Exploration of the South Pacific* (ed. Wooster, W. S.) 183–193 (National Academy of Science, Washington DC, 1970).
- Jahnke, R. A. *Rev. Geophys.* **28**, 381–398 (1990).
- Dugdale, R. C. & Goering, J. J. *Limnol. Oceanogr.* **12**, 196–206 (1967).

- Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677–680 (1979).
- Platt, T. & Harrison, W. G. *Nature* **318**, 55–58 (1985).
- Gorzelska, K. & Galloway, J. N. *Global. biogeochem. Cycles* **4**, 309–333 (1990).
- Glover, H. E., Prezelin, B. B., Campbell, L., Wyman, M. & Garside, C. *Nature* **331**, 161–163 (1988).
- Okubo, A. *Deep-Sea Res.* **18**, 789–802 (1971).
- Willey, J. D. & Cahoon, L. B. *Mar. Chem.* **34**, 63–75 (1991).
- Woodward, E. M. S. & Owens, N. J. P. *Neth. J. Sea Res.* **25**, 57–63 (1990).

ACKNOWLEDGEMENTS. We thank A. Gaudry and the staff of the Bermuda Biological Station for coordinating precipitation sampling on Amsterdam Island and Bermuda, respectively, and W. Keene, J. Maben, J. Montag and C. Reinhart for doing the chemical analyses. This research forms part of Research Project II of the Plymouth Marine Laboratory (a component of the UK NERC) and is a contribution to the Bermuda Biological Station for Research Inc., the Western Atlantic Ocean Experiment and the Global Precipitation Chemistry Project. This analysis was supported by WMO, UNEP and UNESCO as part of a GESAMP workshop.

Rapid racemization of aspartic acid in mollusc shells and potential for dating over recent centuries

Glenn A. Goodfriend*

Department of Environmental Sciences and Energy Research, Weizmann Institute of Science, 76100 Rehovot, Israel and Geophysical Laboratory, Carnegie Institution of Washington, 5251 Broad Branch Road, N.W., Washington, DC 20015, USA

DATING of deposits and materials less than 350 years old is hindered by the very poor time resolution of the radiocarbon method over this period^{1,2}. Fluctuations in atmospheric ¹⁴C levels result in the existence of several possible calendric ages for any given radiocarbon age for terrestrial samples. In the marine record, these fluctuations are dampened. However, only a small change in marine radiocarbon ages occurs over this period (from AD 1700 to 1950, radiocarbon ages become only 100 yr younger³). Consequently, age resolution is poor. Furthermore, there is an uncertainty in the amount of the correction for the reservoir effect, the apparent radiocarbon age of modern marine carbon (typically ~400 yr), which reflects the average residence time of carbon in the oceans. Amino-acid racemization/epimerization analysis has been used primarily for dating samples older than the limit of radiocarbon dating (40,000–50,000 yr BP)^{3,4}. Here I show that aspartic acid (Asp) in mollusc shells has racemized particularly rapidly over the past few centuries. Analyses done on land snails from deposits in the Negev Desert, eolianites in Madeira, cave deposits in Jamaica and from museum collections indicate racemization rates of 2–5% per century. Asp racemization thus provides a means of dating recent faunal assemblages of molluscs and other biogenic carbonates, as well as recent terrestrial, fluvial and marine sedimentary sequences that contain molluscs.

Racemization of Asp has been studied mainly in bones and teeth⁵. In living humans, teeth show a relatively rapid rate of racemization of aspartic acid, Asp (1% per 13 yr, or 8% per 100 yr, in dentine⁶). But at ambient temperatures, this rate would be much slower (0.56% per 100 yr at 21 °C; ref. 7). Rapid Asp racemization has been observed in parchments, in which the collagen has been altered to gelatin⁸. Asp racemization in mollusc shells has received little attention. Until recently, only Late Pleistocene samples from marine sediments of the west coast of the United States had been studied⁹, and these were not found

TABLE 1 Estimated aspartic acid (Asp) racemization rates for land snail shells

Sample series	Epimerization rate (% per 100 year)	Ratio of rates of racemization to epimerization*	Asp racemization rate (% per 100 yr)
Negev	0.238†	12.0	2.9*
Jamaica	0.217‡	11.1	2.4*
Madeira	0.148‡	13.1	1.9*
Ottoman RR§	—	—	3.9
Museum	—	—	5.3

* Average rate for samples with D/L ≤ 0.20.

† Based on radiocarbon-dated samples; see Fig. 1a.

‡ Data in Fig. 1. § See text. || Data in Fig. 2b.

to have a high rate of Asp racemization. Younger mollusc shells (Holocene land snails from the Negev Desert, southern Israel) have recently been studied¹⁰. This series of radiocarbon-dated samples suggests a very rapid rate of racemization of Asp up to a D/L ratio of ~0.3 at 1,000 yr BP, then a slower rate in samples from 1,000 to 10,000 yr BP. The time resolution of radiocarbon in the past 1,000 years is, however, poor for these shells, as (1) ages cannot be resolved for the past 350 yr, and (2) the correction for the age anomaly (resulting from ingestion of old carbonates¹¹) adds an uncertainty of ~200 yr to the analytical error of the radiocarbon ages¹².

Here I bring together several different lines of evidence to show that very rapid initial racemization of Asp is a widespread phenomenon in snail shells. I document the pattern of Asp racemization in the Negev land snails (*Trochoidea seetzeni*) in greater detail, and document those in two other time series of land snail shells: one from eolianite deposits in eastern Madeira (*Caseolus bowdichianus* and *Theba pisana*) and the other from sediments in a cave in central Jamaica. The Jamaican series includes both pulmonate (*Pleurodonte* spp.) and prosobranch (*Annularia pulchrum*) land snails. (The prosobranchs are more closely related to marine prosobranchs than to pulmonate land snails.) I also obtain patterns of initial racemization of Asp in heating experiments on live-collected Negev land snail shells and the degree of racemization of known-age snail shells, excavated from the embankment of the Ottoman railroad (built in 1916; ref. 13) in the Negev Desert. Finally, I analyse a 110-year series of live-collected, known-age shells of *Triodopsis*, a North American land snail, from the National Museum of Natural History and my own recent collections.

Because the radiocarbon timescale provides poor age resolution for land snail samples in the late Holocene, I evaluate the time trend of D/L Asp values of the fossil samples against a different measure of age; the D-alloisoleucine/L-isoleucine (or A/I) epimer values. A/I values increase linearly with radiocarbon ages and show a very strong correlation (0.94) with radiocarbon ages over the last 10,000 years in the Negev land snails¹⁰. It can be assumed that this strong correlation and linearity also exists for the late Holocene, when these cannot be evaluated directly because of the imprecision of radiocarbon dates. The A/I values are converted to a timescale on the basis of calibration dates which are within the time range when radiocarbon dates are more precise.

In all three sets of fossil samples (Fig. 1) a similar pattern is found: the initial rate of racemization of Asp is very rapid but slows down progressively with increasing D/L. I estimated the average initial rates of Asp racemization in the fossil sample series (for samples with $D/L \leq 0.20$) by calculating the rate of Asp racemization relative to A/I epimerization in samples with $D/L \leq 0.20$ (by simple linear regression of D/L on A/I), and the average Holocene rate of A/I epimerization based on calibration of A/I values against radiocarbon ages. The ratio of the initial rate of racemization to rate of epimerization is similar (11–13) for all three fossil series (Table 1). Epimerization rates were similar for the Negev and slower for Madeira (Table 1). Together, these estimates indicate that average Asp racemization rates vary from 2.9% per 100 yr for the Negev samples to 1.9% per 100 yr for the Madeiran samples. Because

the trend of D/L with time (A/I) is curved, the initial rate of racemization will be higher than this average, whereas the later rate (nearer $D/L = 0.20$) will be slower than the average. Results from heating experiments (Fig. 2a) show that this pattern of racemization (decreasing rate with increasing D/L value) is reproduced under controlled constant temperature conditions in the laboratory.

Analyses of *Trochoidea seetzeni* shells from the embankment of the Ottoman railroad in the Negev (6 km south of junction with the Dimona Road, south of Beer Sheva) provide an independent means of determining the rate of Asp racemization in very recent shells. The embankment was constructed by piling up the local silty sediments. In the process, the shells of live or freshly dead snails lying on the surface, would be incorporated, along with any fossil material present in the sediments. The D/L Asp values of the 10 individuals analysed range from 0.081 to 0.432, indicating the presence of redeposited shells. But the three specimens with the lowest values (0.080, 0.084; 0.080, 0.082; 0.082, 0.091; duplicate preparations of each individual are separated by commas) show consistency within analytical error (2% of the mean) and are taken to represent the shells that were modern at the time of construction of the railroad (1916). From the mean D/L values of these three shells (0.084) and the D/L value of live-collected *T. seetzeni* (0.055; ref. 14), a net racemization rate of 2.9% in 75 yr (or 3.9% per 100 yr) was calculated. As expected, this rate is higher than the average rate (2.9% per 100 yr) seen in the Negev sample series with $D/L \leq 0.20$, because of the curvilinear (convex upward) trend of D/L values with time. This time series encompasses the past 500 yr, according to the calibration of A/I against radiocarbon ages.

Another check on the relation of age to D/L Asp values in young shell material comes from analysis of a series of museum samples of *Triodopsis*. These show a high racemization rate of 5.3% per 100 yr (Fig. 2b; Table 1) and a trend of increasing D/L Asp value with increasing age, with a scatter of ± 15 yr (square root of the mean square error) about the trend line. This scatter, which is beyond the analytical error, might be due to racemization that may have been induced by heating of the shells to dry them after the soft parts were removed, or to differences in the relative thickness and rates of racemization of the different shell microstructural units.

The differences in racemization rates in different sets of samples probably relate primarily to temperature differences, with museum temperatures being the warmest and the tem-

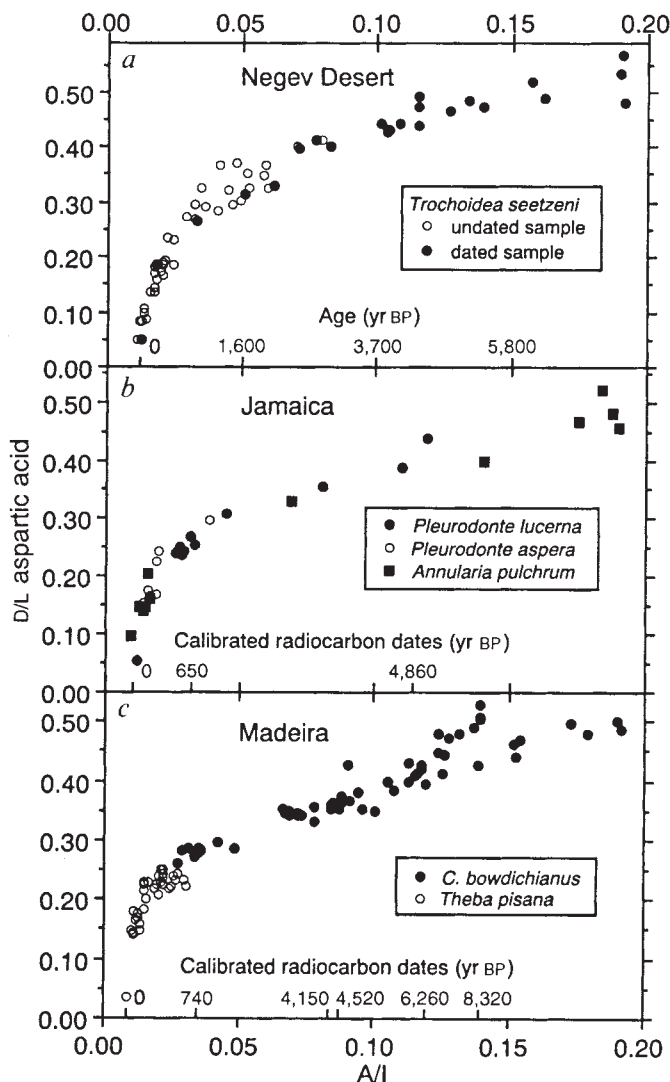


FIG. 1 D/L aspartic acid against D-alloisoleucine/L-isoleucine (A/I) values in land snails. *a*, *Trochoidea seetzeni* from the Negev Desert. The age equivalents of the A/I values are indicated on the inner label of the abscissa. These are calculated from the 30 samples in ref. 10 having radiocarbon ages of <8,000 yr BP (all those within the range of the dendrochronology calibration) and one modern sample. *b*, Three species excavated from a cave at Coco Ree, St Catherine, Jamaica. Calibrated radiocarbon dates and A/I values of two samples of *Pleurodonte lucerna* shells²⁴ are indicated. *c*, Two species from the eolianite deposits on the São Lorenzo Peninsula, eastern Madeira. Calibrated radiocarbon dates and A/I values of five samples of *Caseolus bowdichianus* shells (G.A.G. et al., manuscript in preparation) are indicated. *C. bowdichianus* is extinct on Madeira. Measurements were carried out by gas chromatographic analysis of the *N*-trifluoroacetyl isopropyl ester derivatives^{10,22} of the amino acids in hydrolysates of cleaned shells. Analyses of the fossil material were done on aliquots of individual shells as well as of preparations of many shells ($n=10-35$) used for radiocarbon dating. The individual shells were generally measured once, whereas several measurements were made on aliquots of the bulk preparations for radiocarbon dating¹⁰. For the radiocarbon-dated samples, errors average 3.5% of the A/I value and 1.4% of the D/L Asp value. Errors for single measurements of the other specimens average 5% for A/I (10% in the range of 0.01–0.02) and 2% for D/L Asp. All A/I measurements were calibrated against A/I standard mixtures; these indicate a systematic bias that required multiplication of peak area ratios by 0.87 to obtain true values. All radiocarbon dates are reported corrected for isotopic fractionation and have been calibrated using the curves in ref. 23. The Negev and Madeira snail dates were corrected for age anomalies¹² before calibration.

peratures on Madeira the coolest. Variation in racemization rates amongst species is probably another factor.

A very high initial rate of racemization of Asp thus seems to be a widespread phenomenon in snail shells. It occurs also in marine molluscs (G.A.G., D. W. Von Endt and P. E. Hare, unpublished data) and stony corals (G.A.G., P. E. Hare and E. M. Druffel, in preparation) and apparently also in ostrich egg shells⁷. It therefore seems to be a general phenomenon of amino acids in the organic matrix of biogenic calcium carbonates.

Racemization/epimerization reactions are well known to be sensitive to temperature. For isoleucine epimerization in molluscs, the rate increases by ~15% per °C at typical ambient temperatures. Asp racemization in molluscs has a similar (or marginally lower) temperature dependence¹⁴. Samples used to calibrate racemization/epimerization rates should ideally come from the same site as the samples to be dated, to minimize possible temperature differences. Rates of racemization will also differ amongst different taxa¹⁵. These relative rates can be determined from separate calibrations, from heating experiments²⁴, or by comparing D/L values between species in a series of deposits^{15,16}.

The rapid rate of racemization of Asp can be expected to provide good time resolution for samples dating from the past three centuries, for which radiocarbon analysis is of little use. Epimerization/racemization analysis has been used for several relative dating applications, most commonly for correlation of stratigraphic units (for example, ref. 4). It has also been used

to identify and sort mixed-age mollusc assemblages^{17–20}. In addition, epimerization analysis of known-age material has been used for palaeotemperature analysis^{19,21}. Asp racemization could be used in each of these applications on a much more recent timescale than had previously been envisaged, and also opens up new dating possibilities, such as the determination of age of museum specimens. The latter application may be of particular use to biogeographers wishing to know whether specimens in collections were alive when collected or whether they represent dead-collected material, attesting to an earlier occurrence of a species at a site. Other possible applications include dating of recent deposits such as river terraces, marine and lake sediments, dunes, storm accumulations on beaches, or historical-age archaeological sites; the study of taphonomic processes in recent environments; and determination of longevity of molluscs. Asp racemization may also provide better time resolution than does epimerization for deposits in cold environments (polar, high altitude or deep sea). □

Received 21 October 1991; accepted 6 April 1992.

1. Stuiver, M. & Pearson, G. W. *Radiocarbon* **28**, 805–838 (1986).
2. Stuiver, M., Pearson, G. W. & Braziunas, T. F. *Radiocarbon* **28**, 980–1021 (1986).
3. Bada, J. L. A. *Rev. Earth planet. Sci.* **13**, 241–268 (1985).
4. Wehmiller, J. F. *Quat. Sci. Rev.* **1**, 83–120 (1982).
5. Bada, J. L. *Interdiscip. Sci. Rev.* **7**, 30–46 (1982).
6. Helfman, P. M. & Bada, J. L. *Nature* **262**, 279–280 (1976).
7. Goodfriend, G. A., von Endt, D. W. & Hare, P. E. A. *Rep. Director Geophys. Lab. Carnegie Inst. Wash.* 1990–1991, 172–176 (1991).
8. Weiner, S., Kustanovich, Z., Gil-av, E. & Traub, W. *Nature* **287**, 820–823 (1980).
9. Kvendovden, K. A., Blunt, D. J. & Clifton, H. E. *Geochim. cosmochim. Acta* **43**, 1505–1520 (1979).
10. Goodfriend, G. A. *Geochim. cosmochim. Acta* **55**, 293–302 (1991).
11. Goodfriend, G. A. & Stipp, J. J. *Geology* **11**, 575–577 (1983).
12. Goodfriend, G. A. *Radiocarbon* **29**, 159–167 (1987).
13. Avitsur, S. *Proc. Israeli Geogr. Soc. Ann. Meeting* (in Hebrew) 11–17 (Department of Geography, Ben Gurion Univ., Beer Sheva, 1977).
14. Goodfriend, G. A. & Meyer, V. R. *Geochim. cosmochim. Acta* **55**, 3355–3367 (1991).
15. Lajoie, K. R., Wehmiller, J. F. & Kennedy, G. L. in *Biogeochemistry of Amino Acids* (eds Hare, P. E., Hoering, T. C. & King, K. Jr) 305–340 (Wiley, New York, 1980).
16. Kaufman, D. S. *Geol. Soc. Am. Bull.* **104**, 40–52 (1992).
17. Miller, G. H. & Hare, P. E. in *Biogeochemistry of Amino Acids* (eds Hare, P. E., Hoering, T. C. & King, K. Jr) 415–443 (Wiley, New York, 1980).
18. Goodfriend, G. A. *Quat. Res.* **28**, 374–392 (1987).
19. Goodfriend, G. A. & Mitterer, R. M. *Palaeogeogr. Palaeoecol. Palaeoclim.* **63**, 293–311 (1988).
20. Goodfriend, G. A. *Radiocarbon* **31**, 1041–1047 (1989).
21. Miller, G. H. et al. *Nature* **326**, 593–595 (1987).
22. Abdalla, S., Bayer, E. & Frank, H. *Chromatographia* **23**, 83–85 (1987).
23. Stuiver, M. & Kra, R. (eds) *Radiocarbon* **28**(2B) (1986).
24. Goodfriend, G. A. & Mitterer, R. M. *Geol. Soc. Am. Bull.* (in the press).

ACKNOWLEDGEMENTS. Material from the National Museum of Natural History was provided by R. Herschler and P. Greenhall. Material from the cave at Coco Ree, Jamaica, was excavated by T. Patton and provided by R. Portell of the Florida State Museum. L. M. Cook and R. A. D. Cameron assisted with excavation of snails from Madeira. E. Negreanu assisted with the enantiomeric analysis. The Weizmann Institute of Science radiocarbon dates were provided by I. Carmi. Comments on the manuscript were provided by P. E. Hare and D. von Endt. Field work in Madeira was supported by the National Geographic Society, and analysis of the Negev material by the Ministry of Energy and Infrastructure of Israel.

Linking short-timescale deformation to long-timescale tectonics

Didier Sornette* & Jean Virieux†

* Laboratoire de Physique de la Matière Condensée, CNRS URA 190, Université de Nice-Sophia Antipolis, Parc Valrose, 06108 Nice Cedex 2, France

† Institut de Géodynamique, CNRS URA 1279, Université de Nice-Sophia Antipolis, Sophia Antipolis, 1 avenue Albert Einstein, 06560 Valbonne, France

ONE of the great successes of plate tectonics has been to provide a link between deformations of the lithosphere observed over widely varying timescales. For example, the slip between two plates as measured from young magnetic anomalies (and hence averaged over the past 1–2 Myr) has been found^{1–3} to be compatible with estimates obtained from slip during recent earthquakes, despite the small time interval covered by the earthquake observations.

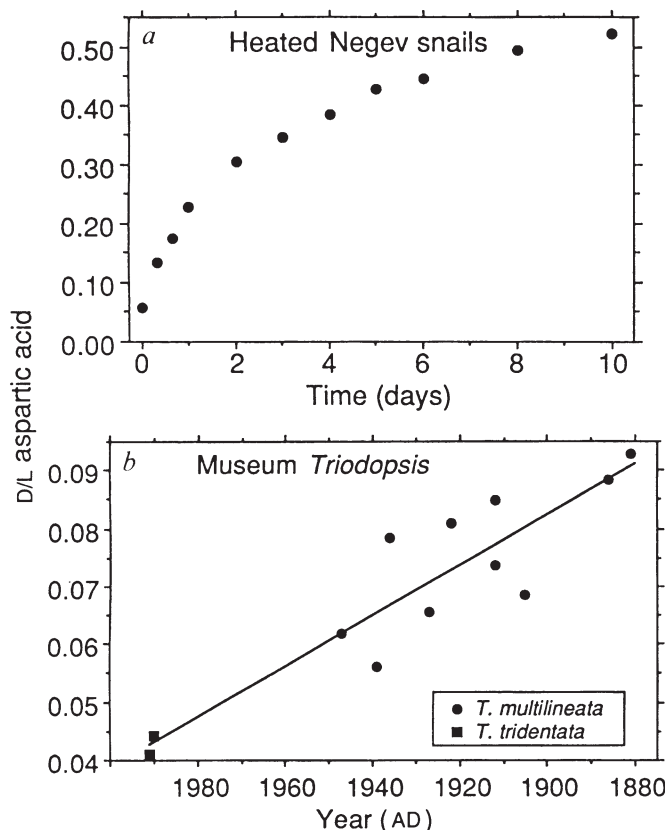


FIG. 2 D/L Asp values in land snail shells of known age. *a*, Time course of racemization of modern shells of *Trochoidea seetzeni* heated at 106.5 °C. See ref. 14 for experimental details. Data for even-number days are taken from ref. 14. *b*, shells of the land snails *Triodopsis multilineata* (from the collection of the National Museum of Natural History) and *T. tridentata* (collected in Washington, DC) in relation to the date of collection of the snails. The correlation coefficient is 0.91. Part of the data come from ref. 7. Duplicate analyses of single preparation of pieces from three individual shells (with periostracum removed) were carried out. Errors average 1.4% for the D/L Asp values.

Here we start from the idea^{4,5} that geological deformations are the long-term cumulative trace of short-term processes such as earthquakes, so that the latter can be described as a high-frequency 'noise' for the former. We return to a theoretical framework⁶ introduced to model large-scale and long-term tectonics, based on a nonlinear diffusion equation, but here explicitly associate the 'noise' term with earthquakes. The observed power-law relationship between earthquake frequency and magnitude, together with an assumption that strain is on average scale-independent, leads to the prediction that the largest earthquakes will have a Gutenberg-Richter 'b-value' of 1.5, as has recently been reported for large ($M_w \geq 7$) earthquakes worldwide⁷.

We start, as in ref. 6, with the tensorial nonlinear Langevin equation: $\partial \varepsilon_{ij} / \partial t = A \partial[(S)_{ij} \partial \varepsilon_{kl}] / \partial x_k + B \partial[(S)_{il} \partial(\varepsilon_{km} \varepsilon_{ml})] / \partial x_j + \partial \eta_{ij}(x, t) / \partial x_k + \eta_{ij}(x, t)$. Here ε is the strain, $\eta_{ij}(x, t)$ is the strain field and $(S)_{ij}$ is the equilibrium tensor field within the plate. The first term of the right-hand side corresponds to strain diffusion, the second represents the nonlinear rheological behaviour of the lithosphere, and the noise term $\eta_{ij}(x, t)$ accounts for the fact that tectonic deformations are locally irregular. The general approach has been justified by the argument that the lithosphere is in a 'critical' state of marginal mechanical stability, characterized by long-range correlations and power laws^{8,9}. In ref. 6, the noise term $\eta_{ij}(x, t)$ was argued to take into account the irregularities of the strain evolution due to local heterogeneities of the lithosphere, in composition, faulting, thickness, coupling with lower layers and so on, as well as the fluctuations of the deformations applied at the borders of the tectonic plates. The strain fluctuations were thus thought to originate mostly from the heterogeneous structure of the lithosphere.

Here, we point out that the noise term $\eta_{ij}(x, t)$ is constrained by the very nature of the Langevin mathematical framework and in fact represents the source of the strain-field fluctuations around its long-term trend. A Langevin equation contains two main contributions: (1) a continuous one describing the long-term average evolution and (2) short-time fluctuations around the average evolution¹⁰. For instance, the Langevin equation describing the motion of a brownian soot particle falling through a gas contains both the average effect of the colliding molecules, which is a smooth drag term proportional to the particle velocity, and the instantaneous effect of these collisions, which is a fluctuating force at times comparable to the molecule-particle collision time and causes the erratic motion of the particle. In the case of tectonic deformations, the driving forces that cause plate motion are the analogue of gravity. The lithosphere heterogeneity, fault roughness and the resulting stick-slip friction rheology cause the earthquakes, which are the analogue of the molecular collisions. Averaged over a long time, a brownian particle seems to be falling through a viscous medium; fault motions and continental deformations seem to be controlled by a smooth friction. We thus argue that the noise term $\eta_{ij}(x, t)$ should describe the effects of short-term deviations from the long-term trend. This noise term must describe the effect of all events on the fluctuations around the long-term trend occurring at timescales much smaller than that at which averaging over the strain ε_{ij} is valid. Therefore, $\eta_{ij}(x, t)$ must describe the effect of the various heterogeneous modes of deformations at short times, such as creep, silent and slow earthquakes¹¹, as well as genuine earthquakes of all sizes¹².

But there is a key difference between plate tectonics and brownian motion: whereas the strength of molecular collisions can be described in terms of a mean and standard deviation, earthquakes are known to follow a power-law size distribution. Therefore, the noise η should be characterized by a probability distribution $P(\eta)$ which cannot be taken to be gaussian, as is usually done in Langevin equations. Consider for instance the earthquakes, which account for less than 10% up to nearly 100% of the deformation in different parts of the world. Their energies E (or seismic moments) are distributed according to the Gutenberg-Richter law $\log_{10} N(M_E) = a - bM_E$, which gives the num-

ber of earthquakes of magnitude smaller than M_E per unit time and per unit surface. Using the standard definition¹³ of the Kanamori magnitude $M_E = (1/c) \log_{10} E - d$, one thus obtains

$$N(E) dE \approx E^{-1-b/c} dE \quad (1)$$

where a and d are two normalizing constants. The phenomenological relation¹⁴ between seismic moment E and size ℓ of the rupture zone associated with a given earthquake is

$$E \propto \ell^\alpha, \quad (2)$$

where $\alpha = 3$ for magnitude $M_w \leq 7$ and is argued to cross over to $\alpha = 2$ for larger magnitudes¹⁴. Combining this with an assumed linear relation¹⁵ between deformation ε (mean slip) and fault rupture size ℓ yields the following probability distribution of the noise η

$$P(\eta) d\eta \approx \eta^{-1-\mu} d\eta \quad (3)$$

with $\mu = \alpha b/c$. The noise is thus distributed according to a power law and shows larger fluctuations than does a well-behaved gaussian. To address the Langevin equation, we must consider all the different contributions to the noise term. Some will give bounded distributions or decay faster than any power law (exponential or gaussian distributions). But it is the contribution with the slowest decaying tail (smallest exponent μ) that controls the nature of the long-range spatio-temporal correlations of the coarse-grained strain field, because it yields the largest fluctuations (see below).

The noise must be characterized by its spatio-temporal correlations. As reliable information is lacking, we assume here that the noise-noise correlations in space and time are short-range, that is, they decay faster than a power law. For instance, we can take the noise to be delta-correlated in space and time as defined in ref. 6. Any other short-range correlations will not modify the results derived below. In future, it may be possible to consider long-range correlations decaying as a power law of time and/or distance¹⁶ and incorporate them in the theory.

If we ignore any long-range noise correlation and the power-law distribution (equation (3)), the Langevin equation can be solved using the dynamical renormalization group¹⁶, and the general solution takes the homogeneous scaling form

$$\varepsilon_{ij}(\ell, \ell_\perp, t) = B^\chi \varepsilon_{ij}(\ell/B, \ell_\perp/B^\zeta, t/B^z) \quad (4)$$

where B is an arbitrary scaling factor. The exponents χ , ζ and z , which define how the strain field scales with space and time, are given exactly by^{6,17} $z = 6/(7-d)(3/2$ in three dimensions), $\chi = (1-d)/(7-d)(-1/2$ in three dimensions) and $\zeta = 3/(7-d)(3/4$ in three dimensions).

If we take the power-law nature of the noise term (equation (3)) into account and assume that the strain is on average scale-independent, the values of χ , ζ and z are modified. As no formal theoretical method is available at present to tackle this case, we develop scaling arguments in the spirit of ref. 18. Although they are not rigorous, these have the advantage of physical clarity. We first simplify the tensorial Langevin equation into an equation in terms of a scalar field ε which describes for instance the second invariant $(\sum_{ij} \varepsilon_{ij} \varepsilon_{ij})^{1/2}$ of the strain tensor. The tensorial character is kept only through the existence of anisotropy, thus defining a main direction and its transverse (denoted by $\perp$). The corresponding Langevin equation then reads

$$\partial \varepsilon / \partial t = D_\parallel \partial^2 \varepsilon / \partial x_\parallel^2 + D_\perp \partial^2 \varepsilon / \partial x_\perp^2 - (\lambda/2) \partial(\varepsilon^2) / \partial x_\parallel + \eta(x, t) \quad (5)$$

This equation is anisotropic ($D_\parallel \neq D_\perp$) and contains the leading nonlinear term $\partial \varepsilon^2 / \partial x_\parallel$ which is allowed in the presence of this anisotropy^{6,17}. Following ref. 18, we assume that at long times $t \gg t_\ell$, and averaged over length scales ℓ , the typical magnitude of the fluctuation of the strain amplitude is ε_ℓ and that these fluctuations last for times of the order t_ℓ . Using the two relevant length scales ℓ and $\ell_\perp \approx \ell^\zeta$ of the problem (where the exponent ζ is to be determined), the various terms in equation (5) can be

estimated as

$$\langle |\partial \varepsilon / \partial t| \rangle_\ell \approx \varepsilon_\ell / t_\ell, \quad D_\parallel \partial^2 \varepsilon / \partial x_\parallel^2 \approx D_\parallel \varepsilon_\ell / \ell^2, \quad (6)$$

$$D_\perp \partial^2 \varepsilon / \partial x_\perp^2 \approx D_\perp \varepsilon_\ell / \ell_\perp^2, \quad (\lambda/2) \partial \varepsilon^2 / \partial x_\parallel \approx \lambda \varepsilon_\ell^2 / \ell$$

We first assume that the nonlinear term $(\lambda/2) \partial \varepsilon^2 / \partial x_\parallel$ dominates the two diffusion terms in the right-hand side of equation (5). Equating $\langle |\partial \varepsilon / \partial t| \rangle_\ell$ with the nonlinear term then implies that a typical fluctuation last for a time $t_\ell \approx \ell / [\lambda \varepsilon_\ell]$; if equation (4) is assumed to hold true, this gives

$$\chi = 1 - z \quad (7)$$

with $\varepsilon_\ell \approx \ell^\chi$ and $t_\ell \approx \ell^z$. Equating $\langle |\partial \varepsilon / \partial t| \rangle_\ell$ with the dominant diffusion term yields

$$z = 2\zeta \quad (8)$$

We need a third relation to determine the three exponents z , χ and ζ . If the noise η was distributed according to a well-behaved gaussian, for instance, the central limit theorem would be valid and the typical order of magnitude of the noise term η_ℓ at scale ℓ would be $\eta_\ell \approx (D/\ell^{d-1})^{1/2}$. Equating $\langle |\partial \varepsilon / \partial t| \rangle_\ell$ with this estimate for the noise amplitude would yield $z = 1 + \zeta(d-1)/3$. Together with equations (7) and (8), this recovers the exponents derived in refs 6 and 17 and quoted below equation 4.

This last relation $z = 1 + \zeta(d-1)/3$ must be modified for power-law noise, as very large noise fluctuations can appear in this case and the central limit theorem no longer holds. Physically, what happens in this case is that rare exceptional noise events lead to large strain fluctuations. Estimates of the noise in terms of averages do not reflect this fact. We must instead estimate the size of the largest relevant noise fluctuation. This is done using standard extreme probability theory¹⁹. We further recognize that, at large times, the strain evolution becomes very anisotropic since $\ell \gg \ell_\perp \approx \ell^\zeta$ with $\zeta < 1$. It is therefore reasonable to assume that exceptional noise fluctuations occur mostly in the longitudinal direction. Geologically speaking, only the few largest faults, on which the largest earthquakes occur, dominate. During a time t_ℓ , there are roughly $t_\ell \ell$ independent noise realizations. The typical maximum noise $\eta_{\max}$ encountered in such $t_\ell \ell$ trials is such that $(t_\ell \ell) \int_{\eta_{\max}}^{\infty} P(\eta) d\eta \approx 1$

$$\eta_{\max} \approx (t_\ell \ell)^{1/\mu} \quad (9)$$

Note that the power law (equation (3)) for $P(\eta)$ implies a strong power-law dependence of $\eta_{\max}$ as a function of ℓ , in contrast to the extremely weak logarithmic dependence if $P(\eta)$ was gaussian, for which the central limit theorem would hold.

Physically, the noise term tends to increase the strain fluctuations. These fluctuations are resisted by a 'rigidity' of the strain stemming from its diffusive behaviour, as one can see by examination of the linear diffusion equation with gaussian noise, $\partial \varepsilon / \partial t = D \nabla^2 \varepsilon + \eta(r, t)$. In this case the r.m.s. fluctuation $\Delta \varepsilon$ is, for asymptotically large times, $\Delta \varepsilon \approx \log \ell$ in two dimensions and constant in three dimensions in a system of size ℓ (ref. 20), as compared to the general 'random walk' result $\Delta \varepsilon \approx \ell^{1/2}$ independent of ℓ in all dimensions for $D = 0$. A non-vanishing diffusion coefficient D , representing a kind of viscosity on the strain field, imposes smooth enough strain fluctuations. From this line of reasoning, we can estimate the 'rigidity' of the strain stemming from its diffusive behaviour by introducing the 'entropy' $\sim \ell^2 / t_\ell$ lost in spanning the scale ℓ in time t_ℓ (ref. 21).

The final step of the reasoning is to balance this 'entropy' $\sim \ell^2 / t_\ell$ with the 'energy' $\eta_{\max}$ gained over this volume: $\ell^2 / t_\ell \approx \eta_{\max}$. This relation replaces $\langle |\partial \varepsilon / \partial t| \rangle_\ell \approx \eta_\ell$ which held for gaussian noise. Using equation (9) for $\eta_{\max}$, we finally obtain

$$z = (2\mu - 1) / (\mu + 1) \quad (10)$$

Equations (4) and (7-10) provide the scaling of the solution of the Langevin equation. We have confidence in the validity of our reasoning, because it has recently been checked accurately in the context of the growth of an interface described by the KPZ Langevin equation²², using intensive numerical simulations

on a 16K-Connection Machine (R. Bourbonnais, J. Kertész and D. E. Wolf, Preprint HLRZ-KFA Jülich, 1991). These results show how the deformations at small scales control, through the value of the exponent μ , the correlations of the strain field at large scales.

We now argue that an additional self-consistent condition must be implemented: this is that $\chi = 0$. When $\chi = 0$, the typical fluctuations of the strain field $\varepsilon_{ij}(\ell, \ell_\perp, t)$ become independent from the scale ℓ . This ensures complete coherence of the deformation over very large distances. The condition is also born out from observations^{14,15} showing that the average slip Δu over a fault of length ℓ is proportional to ℓ , thus defining a strain $\varepsilon \approx \Delta u / \ell$ independent of ℓ . This implies that the strain exponent $\chi = 0$, and that $z = 1$ from equation (7), $\zeta = 1/2$ from equation (8) and

$$\mu \equiv \alpha b / c = 2 \quad (11)$$

from equation (10). This last self-consistent argument makes explicit how large-scale and long-lasting deformations may control the deformation on short timescales; the exponent μ of the probability distribution of the noise η describes the small-scale deformations.

Assuming now that earthquakes provide the largest fluctuations on the short-time deformations, so that the Gutenberg-Richter law yields the noise distribution (3) with the smallest coefficient μ , we deduce that the ratio of the two coefficients b and c involved in the earthquake size distribution of a three-dimensional medium is

$$b/c = 2/\alpha \quad (12)$$

Because of the lack of reliable data, we will assume for the rest of the discussion that the coefficient c is fixed at $c = 1.5$. Then, if we take $\alpha = 3$ as the accepted value for 'small' earthquakes ($M_w \leq 7$), we obtain $b \approx 1$. It is interesting to note that 'small' ($M_w \leq 7$) earthquakes worldwide, for which $\alpha = 3$ is well-founded, do follow the Gutenberg-Richter distribution with b close to one. The value $b = 1$ is also found in dynamical models²³ of small earthquakes in a single large fault. It is clear from our discussion, however, that 'small' ($M_w \leq 7$) earthquakes are irrelevant for the description of long-time and large-scale tectonic deformations. Indeed, their contribution to the deformation is negligible compared with that caused by the largest earthquakes. In the other limit $\alpha = 2$ argued to hold asymptotically for very large earthquakes¹⁴, we find $b = c = 1.5$. These predictions are consistent with the results of Pacheco *et al.*⁷, who find $0.8 \leq b \leq 1.2$ for small earthquakes and $b \approx 1.5$ for larger earthquakes for which the rupture dimension is larger than the downdip width of the seismogenic layer. $\square$

Received 29 August 1991; accepted 7 April 1992.

1. Kanamori, H. *J. geophys. Res.* **82**, 2981-2987 (1977).
2. Kanamori, H. & Stewards, G. S. *Phys. Earth planet. Inter.* **11**, 312-332 (1976).
3. Stewards, L. M. & Okal, E. A. *J. geophys. Res.* **88**, 10495-10507 (1983).
4. King, G. C. P., Stein, R. S. & Rundle, J. B. *J. geophys. Res.* **93**, 13307-13318 (1988).
5. Sornette, A. & Sornette, D. *Europhys. Lett.* **9**, 197-202 (1989).
6. Sornette, D., Davy, P. & Sornette, A. *J. geophys. Res.* **95**, 17353-17361 (1990).
7. Pacheco, J. F., Scholz, C. H. & Sykes, L. R. *Nature* **355**, 71-73 (1992).
8. Sornette, D. *Proc. NATO ASI, Spontaneous Formation of Space-Time Structures and Criticality* (eds Riste, T. & Sherrington, D.), 57-106 (Kluwer, Dordrecht, 1991).
9. Scholz, C. H. & Mandelbrot, B. B. (eds) *Fractals in Geophysics* (Birkhäuser, Basel, 1989).
10. Risken, H. *The Fokker-Planck equation*, 2nd edn (Springer, Berlin, 1989).
11. Beroza, G. C. & Jordan, T. H. *J. geophys. Res.* **95**, 2458-2510 (1990).
12. Rundle, J. B. *J. geophys. Res.* **93**, 6237 (1988).
13. Ekström, G. & Diewonski, A. M. *Nature* **332**, 319 (1988).
14. Scholz, C. H. *The mechanics of Earthquakes and Faulting* (Cambridge University Press, 1990).
15. Ranalli, G. *Rheology of the Earth: Deformation and Flow Processes in Geophysics and Geodynamics* (Allen & Unwin, Boston, 1987).
16. Medina, E., Hwa, T., Kardar, M. & Zhang, Y.-C. *Phys. Rev. A* **39**, 3053-3075 (1989).
17. Hwa, T. & Kardar, M. *Phys. Rev. Lett.* **62**, 1813-1816 (1989).
18. Hentschel, H. G. E. & Family, F. *Phys. Rev. Lett.* **66**, 1982-1985 (1991).
19. Galambos, J. *The Asymptotic Theory of Extreme Order Statistics* (J. Wiley, New York, 1978).
20. Edwards, S. F. & Wilkinson, D. R. *Proc. Roy. Soc. A* **381**, 17 (1982).
21. Wax, N. (ed) *Selected Papers on Noise and Stochastic Processes* (Dover, New York, 1954).

22. Kardar, M., Parisi, G. & Zhang, Y. C. *Phys. Rev. Lett.* **56**, 889–892 (1986).
 23. Carlson, J. M. & Langer, J. S. *Phys. Rev. Lett.* **62**, 2632–2635 (1989).

ACKNOWLEDGEMENTS. D.S. acknowledges useful discussions with Y. C. Zhang. This work is partially supported by the Program DBT 1991 of CNRS and the GDR 'Physique des Milieux Hétérogènes Complexes' of CNRS and Ministère de la Recherche et de la Technologie.

New Middle Pleistocene hominid crania from Yunxian in China

Li Tianyuan* & Dennis A. Etler†

* Hubei Institute of Archeology, 67 Donghu Road, Wuchang, Wuhan, Hubei 430077, People's Republic of China

† Laboratory for Human Evolutionary Studies, Department of Anthropology, University of California, Berkeley, California 94720, USA

Two fossil human crania have been found in Middle Pleistocene terrace deposits of the Han River in Yun county (Yunxian), Hubei province, China (Figs 1 and 2)¹. These damaged but relatively complete adult specimens show a mixture of features associated both with *Homo erectus* and with 'archaic *H. sapiens*'. The Yunxian crania (Figs 3 and 4), although crushed and distorted to varying degrees, are unusual in having major elements of the basicranium, palate, face and cranial vault preserved together. The specimens reveal many details of facial and basicranial anatomy rarely seen in hominid crania of comparable antiquity. Moreover, they are the most complete crania of such great age discovered on the Asian mainland. They consequently throw new light on Middle Pleistocene hominid diversity and the relationships among regionally disparate Middle Pleistocene hominids.

The two crania (EV 9001 and EV 9002) were discovered *in situ*, encased in a hard calcareous matrix. They are both extremely large, fully adult, presumably male specimens. Taking into account significant differences in their states of preservation the two crania appear to be generally similar in their overall morphology. Cranial vault measurements of EV 9001 are hard, if not impossible, to obtain owing to damage to and distortion of the specimen. EV 9002 is better preserved but still considerably fractured and distorted, making reliable measurement a problem. It is sufficiently well preserved, however, to allow for rough estimates of greatest cranial length (glabella–opisthocranium length ~217), greatest cranial breadth (euryon–euryon breadth ~170) and basibregmatic height (~119). These measurements are presented merely to give an indication of the great length, breadth and overall massiveness of the crania under

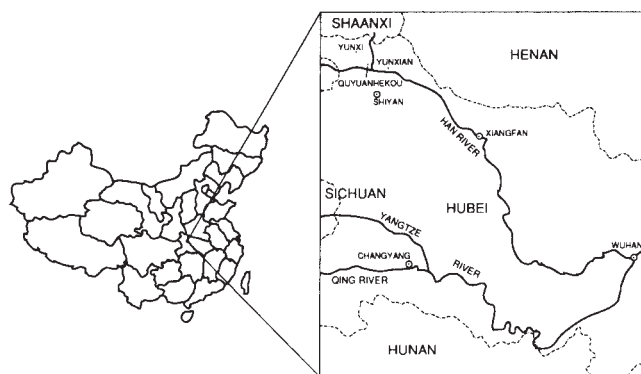


FIG. 1 Maps showing geographic location of Yunxian hominid site at Quyuankou. The hominid crania, found in May 1989 and June 1990, come from a locality near Qingqu village, Mituosi township at the confluence of the Quyuankou and Han Rivers called Quyuankou. The locality is situated on an east/west trending ridge on top of a high terrace of the Han River.

discussion and are not meant to be used as comparative metric data.

Before the discovery of EV 9001 and EV 9002, fossils of *H. erectus* recovered in China consisted solely of skull caps, craniofacial fragments, mandibles, isolated teeth and fragmentary post-crania excavated at Zhoukoudian^{2–5}, Lantian⁶, Hexian⁷ and a number of other localities⁸. Somewhat younger relatively complete crania, attributed to an early form of *H. sapiens*, are known from Dali⁹, Yingkou (Jinniushan)¹⁰ and Maba¹¹. The new crania from Yunxian are thus important comparative material which will help clarify our understanding of the nature of human variability in east Asia during the Middle and early Late Pleistocene. They will also aid in establishing the relationship of Asian to non-Asian Middle and early Late Pleistocene hominids.

The Yunxian crania show features of the mid-face common to non-neanderthal late archaic¹² and early modern *H. sapiens* (for example, the face is flattened and orthognathic with moderate alveolar prognathism; the maxilla has a distinct canine fossa; the lateral part of the maxilla is oriented coronally and highly angled to the zygomatic; there is a high origin of the zygomatic root; a horizontal inferior zygomaxillary border and a pronounced malar incisure, and so on). These features are associated with the large overall facial dimensions and the robust, posterolaterally oriented supraorbital tori common to Middle Pleistocene hominids from Europe¹³ and Africa¹⁴. The 'sapiens'-like mid-face of the Yunxian crania is also seen in other specimens, both assumed males and females, of east Asian *H. erectus*⁵ and early *H. sapiens*⁹. In contrast, many Middle Pleistocene hominids from Europe and Africa have a different pattern characterized by a pneumatized, obliquely set mid-face and a low origin of the zygomatic root, reminiscent of later-occurring Neanderthals¹⁵. A low origin of the zygomatic root is an ancestral

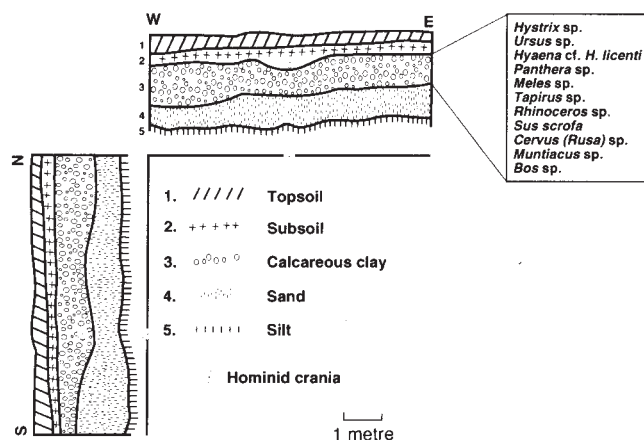


FIG. 2 Stratigraphic profile and mammalian taxa recovered at Quyuankou. The summit of the terrace upon which the locality rests is ~50 m above river level. During terracing operations 12 to 15 m of overlying sediments were removed in the 1970s, leaving behind an overburden of ~1 m of topsoil and subsoil. The hominid and other crania were recovered from calcareous nodules in stratum 3. A lesser quantity of fossil material is found in stratum 4. The sections represented are W/E and N/S trending profiles. They show the position of the hominid crania relative to each other. The profiles join at the upper left-hand corner. The hominid crania were vertically separated by 0.4 m and horizontally separated by 3.3 m. Mammalian fossils associated with the hominids include relatively complete crania, limb bones and isolated teeth of more than ten taxa. The recovered taxa are common elements of the Middle Pleistocene *Stegodon–Ailuropoda* fauna of central and southern China. Twenty-four lithic artefacts were collected from the surface in the vicinity of the excavation site, another 23 were collected at the site itself, and 21 were excavated *in situ*. The artefacts include large cores, unifacial and bifacial tools worked on large cobbles, flakes and flake fragments. Raw material is predominantly quartz and quartzite.

TABLE 1 Distribution of craniofacial character states in Middle Pleistocene hominids

	OH9	S.17	Ng	GWL	ZKD	HX	EV 9002	Dali	Pet/BH
1. Cranial vault long and low	++	++	+	++	++	+	++	+	+
2. Parietal walls converge immediately above position of maximum breadth (supramastoid crests)	+	+	+	?	+	+	+	-	-
3. Frontal flat and receding	+	+	+	+	+	+	+	+	+
4. Short, flat parietal	+	+	+	+	+	+	+	(+)	(-)
5. Occipital plane shorter than nuchal plane	(-)	-	+~	?	+	+	+	-	+~
6. Nuchal and occipital planes highly angled so that opisthocranium equals inion	+	++	+	?	++	+	++	-	+~
7. Strong laterally expansive occipital torus	+	++	++	?	++	+	+	-	-
8. Supraorbital torus, straight in anterior and superior view	++	-	++	-	++	+	-	-	-
9. Ectocranial buttresses well expressed (that is, metopic, sagittal and coronal eminences united around bregma)	-	+	++	+	++	+	(-)	+	-
10. True angular torus on parietal	-	+	++	?	++	+	-	+	-
11. Temporal low with flat superior margin	+	+	+	?	+	-	-	-	-
12. Tympanic plate robust, thick and very long with 'club-like' eustachian process	-	+	++	?	++	++	++	-	-
13. Tympanic oriented in coronal plane, highly angled to petrous	+	++	++	?	++	++	-	-	-
14. Fissure separates mastoid from petrosal crest of tympanic	-	+	+	?	+	+	(-)	-	-
15. Sphenoid does not contribute to medial wall of glenoid fossa	(+)	(+)	(-)	?	(+)	?	+	(-)	(-)
16. Styloid process ossified to cranial base	(+)	-	(-)	?	-	-	-	?	+
17. Straight tribasilar bone	+	+	+	?	+	+	(+)	-	-
18. Upper and midface very flat	+	+	?	?	+	?	+	+	+~
19. Lower border of zygomatic process of maxilla retracted (canine fossa may be present)	?	-	?	-	+	?	+	+	-
20. Root of zygomatic process high on maxillary body	?	-	?	-	+	?	+	+	-
21. Short, broad face	?	+	?	?	?	?	+	+	-
22. Well expressed mid-facial pneumatization	?	+	?	?	?	?	-	-	++/+
23. Lower nasal margin not sharply defined	?	(+)	?	?	?	?	(-)	-	-
24. Large interorbital breadth	-	++	+	+	+	+	++	++	+
25. Thick cranial walls due to expansion of inner and outer tables	+	+	+	++	++	+	?	+	+
26. Mastoid and supramastoid crests unit anteriorly	?	+	-	?	+	+	-	-	-
27. Cranial capacity 700-1,000 ml	+	(+)	-	+	+~	(-)	-	-	-
28. Cranial capacity, 1,000-1,300 ml	-	(-)	+	-	+~	(+)	+	+	+

Traits 1-4 are well expressed in EV 9002 and are common characteristics of *H. erectus*. Archaic forms of *H. sapiens*, exemplified by Broken Hill, Petralona and Dali, have an elevated maximum cranial breadth and more vertically implanted cranial side walls. Traits 5-7 are ubiquitous features of *H. erectus* and are strongly expressed in EV 9002. Specimens of archaic *H. sapiens* tend to have an expanded occipital plane, reduced occipital torus and dissociation of opisthocranium and inion. Traits 8-10 are often said to be derived features of Asian *H. erectus*. They are commonly seen in specimens from Zhoukoudian, Sangiran and Ngandong. These traits are either lacking or slightly developed in European and African specimens of archaic *H. sapiens*. They are also poorly expressed in EV 9002. Trait 11 is the usual condition in specimens of *H. erectus*. Specimens of archaic *H. sapiens* as well as some Asian *H. erectus*, however, have an arched parietal margin of the temporal, as does EV 9002. Traits 12-17 typify the basicranial morphology of most specimens of *H. erectus*. Specimens of archaic *H. sapiens* tend to have a much more gracile and modern looking basicranium. EV 9002 retains traits 12, 15, 16 and 17, but is closer to archaic *H. sapiens* in the expression of trait 13 and possibly trait 14. Traits 18-21 are facial features common to EV 9002 and previously known specimens of *H. erectus* and early *H. sapiens* from China. They are also often seen in specimens of late archaic *H. sapiens* from Africa and early modern *H. sapiens* worldwide. These traits are best expressed today in Asian and Asian-derived populations. Trait 22, by contrast, is common to many Middle Pleistocene specimens of archaic *H. sapiens* from Europe and Africa and later occurring Neanderthals. The remaining traits are differentially expressed in Middle Pleistocene and later fossil hominids. Data from refs 5-7, 17-23, 27-32. Symbols and abbreviations: +, present; -, absent; (+), probably present; (-), probably absent; +~, variable; ++, strongly expressed; ?, indeterminate. Ng, Ngandong; GWL, Gongwangling (Lantian); ZKD, Zhoukoudian; HX, Hexian; Pet/BH, Petrolona/Broken Hill.

trait seen also in Plio-Pleistocene hominids from eastern Africa, the late Early Pleistocene Lantian cranium from China⁶ and Sangiran 17 from Java¹⁶ (see Table 1 for details). It seems, therefore, that the facial structure characteristic of non-neanderthal Late Pleistocene *H. sapiens* was widespread in east Asia at a time when European and African hominids possessed a divergent pattern, characterized by a combination of primitive retentions and features derived away from the modern human condition. Among European and African Middle Pleistocene crania only those attributed to assumed females (Steinheim and Ndutu) may approximate a more modern facial structure.

Despite their derived '*sapiens*'-like facial morphology and large cranial dimensions, the Yunxian hominids preserve many features typical of *H. erectus*^{5,17-21} (such as a long, low, sharply angulated cranial vault; low placement of greatest cranial breadth; cranial sides that slope inwards towards the apex of the vault producing a bell-shaped paracoronar cranial profile; a hyper-robust tympanic region; a long and narrow mandibular fossa which lacks appreciable development of an articular tubercle; and various other basicranial features including, perhaps, a poorly flexed basicranial axis (see Table 1 for details)). It is the retention of these plesiomorphic cranial vault and



FIG. 3 *a*, Oblique and *b*, basal view of EV 9001. The facial skeleton of EV 9001 is intact but is skewed to the right and vertically compressed. The vault bones have been severely fractured and flattened but the cranial base is well preserved. The palate is fractured and the contour of the dental arcade is distorted. The dental crowns, other than the left second premolar, have suffered varying degrees of damage. The left central incisor (LI_1) and first molar (LM_1) crowns are missing. The remaining incisors lack appreciable shovelling of the lingual surface. The cheek teeth are expanded by matrix-filled cracks precluding reliable measurement. They are, nonetheless, very large, with a molar size sequence of $M_3 > M_2 > M_1$.



FIG. 4 *a*, Facial view of EV 9002. The cranial vault and facial skeleton of EV 9002 have been fractured and crushed but are in better condition than in EV 9001. The squamous portion of the frontal has been fractured on the left side and the cranial vault is depressed medial and posterior to the fracture. The right side of the cranium has been altered by crushing. The lateral portions of the supraorbital tori have been damaged. Both zygomatica as well as the zygomatic arches are largely lost. The orbits are compressed inferosuperiorly. *b*, Lateral view of EV 9002. The intact portion of the frontal is low, flat and receding. The inclination of the frontal squama (b-g-i) is estimated to be 43° , and the frontal profile angle (m-g-i) is estimated to be 55° . This compares with means of 43° and 59° at Zhoukoudian. The ophryonic region is characterized by a gently concave depression between the brow ridges and the frontal squama, rather than the prominent post-toral sulcus seen in specimens of *H. erectus* from Zhoukoudian⁵ or the steeply graded, continuous squamous-toral junction seen in the Ngandong hominids¹⁹. In this respect the Yunxian crania appear most similar to hominid specimens from Sangiran (particularly S-17), Arago, Petralona and Broken Hill. The zygomatic process of the temporal is robust and positioned away from the temporal squama by a broad sulcus. The auditory meatus is elliptical and recessed below a strong suprameatal tegmen. These features are commonly seen in *H. erectus*⁵. Fossae for the insertion of the nuchal musculature are shallow and poorly defined, unlike the deeply excavated hollows seen in the Ngandong hominids¹⁹ or Neandertals. *c*, Basal view of EV 9002. The basi-cranium is well preserved with the entire hard palate intact. The occipital bone and cranial base are covered by a network of small matrix filled cracks. The foramen magnum, occipital condyles and mastoid processes are intact. The tooth row is incomplete, lacking central incisors and the left canine. The occlusal surfaces of the preserved teeth are somewhat more worn and fractured than in EV 9001. The left M_3 is reduced in size and peg-shaped. Radiographic observation shows an unerupted diminutive right M_3 . As in EV 9001 $M_2 > M_1$. The postcanine teeth have been expanded by matrix-filled cracks which have fractured and widened the occlusal surfaces, and much occlusal surface detail has been lost to wear. The shape of the dental arcade is very similar to Petralona²⁰. The palate is broad. The postcanine tooth rows are straight and diverge posteriorly. *d*, Superior view of EV 9002. Post-orbital constriction is reduced relative to remains of *H. erectus* from Zhoukoudian and Sangiran. The tori do not form a continuous horizontal bony shelf as in specimens of *H. erectus* from Zhoukoudian. They are swept back posterolaterally as in the Broken Hill, Petralona, Sangiran 17 and Dali crania. The tori are thickest medially (18 mm), approaching the thickness seen in the Bodo¹⁴ and Petralona¹³ crania.

basicranial features that characterizes the taxon *H. erectus* and demonstrates that the Yunxian specimens are best placed within it.

By contrast, some traits commonly used to distinguish Asian *H. erectus* from contemporary hominids farther to the west in Europe and Africa^{19–21} are not well expressed in the Yunxian crania (the degree of expression of various ectocranial buttressing features for example), suggesting that such characters are polymorphic within Middle Pleistocene hominids and do not serve to define *H. erectus* as a distinct hominid species. Similar conclusions have been reached by other investigators who have recently analysed the supposed autapomorphic nature of various *H. erectus* character states^{21–23}.

The Yunxian crania also lack some of the specialized features that characterize the sample of *H. erectus* from Zhoukoudian⁵ (for example the form of the supraorbital tori and ophryonic region of the frontal). In terms of their overall cranial vault and basicranial structure, however, they retain many features seen in the Zhoukoudian sample of *H. erectus*. They are demonstrably more primitive than Chinese hominids from Dali and Yingkou (Jinniushan), generally attributed to 'early *H. sapiens*'²⁴ (see Table 1 for details).

Given the common occurrence of a mid-facial anatomy characteristic of Late Pleistocene non-neanderthal *H. sapiens* in Middle Pleistocene hominid crania of both sexes from east Asia and its less frequent, possibly sex-delimited occurrence in Middle Pleistocene Eurafian hominids, it can be argued that this morphology appears first in Asia and subsequently spread westward^{25,26}. It has been suggested elsewhere that the cranial vault and basicranial structure of modern humans makes its debut in Africa during the Early Pleistocene²⁷. The presence of these latter features in Middle Pleistocene hominids from Europe and Africa and their absence in Middle Pleistocene hominids from Asia (including the crania under discussion), has been used to buttress the argument that modern humans are derived from a non-Asian stock²⁸. We, on the contrary, propose that the differential temporal appearance of modern morphologies in regionally disparate hominids attests to the mosaic nature of human evolution. It is likely that Middle Pleistocene human dispersal patterns and the first appearances of modern human character states through space and time are much more complex than previously thought.

The new evidence of fossil hominid variability in east Asia afforded by the Yunxian crania indicates that Middle Pleistocene hominids were highly polytypic and regionally differentiated, perhaps even at the demic level. But the differential distribution of character states associated with *H. sapiens* in regionally disparate Middle Pleistocene human populations suggests that the events leading to the emergence of modern humans were not restricted to one region of the world alone. In addition, the mix of characters in the Yunxian crania demonstrates that the taxon *H. erectus* is founded on a set of ancestral hominid traits and regional polymorphisms. It hence has no meaning in a cladistic framework. In light of these considerations we feel that it is best to view all Middle Pleistocene hominids in a broad perspective as an essential part of one evolving lineage in direct ancestry to modern humans. □

14. Controy, G. C., Jolly, C. J., Cramer, D. & Kalb, J. E. *Nature* **276**, 67–70 (1978).
15. Rak, Y. *J. hum. Evol.* **15**, 151–164 (1986).
16. Wu, X. *Acta Anthropol. Sin.* **9**, 312–321 (1990).
17. Rightmire, G. P. *Evolution of H. erectus* (Cambridge Univ. Press, 1990).
18. Maier, W. & Nkini, A. *Cour. Forsch.-Inst. Senckenberg* **69**, 123–130 (1984).
19. Stringer, C. *Cour. Forsch.-Inst. Senckenberg* **69**, 131–143 (1984).
20. Howell, F. C. in *Evolution in African Mammals* (eds Maglio, V. J. & Cooke, H. B. S.) 154–248 (Harvard Univ. Press, Cambridge, MA, 1978).
21. Hublin, J. J. *Anthropos (Brno)* **23**, 175–187 (1984).
22. Brauer, G. *Acta Anthropol. Sin.* **9**, 350–358 (1990).
23. Kennedy, G. E. *J. hum. Evol.* **20**, 375–412 (1991).
24. Wu, X. *Acta Anthropol. Sin.* **7**, 287–291 (1988).
25. Pope, G. G. *J. hum. Evol.* **21**, 189–213 (1991).
26. Sohn, S. & Wolpoff, M. H. *Acta Anthropol. Sin.* **9**, 359–370 (1990).
27. Clarke, R. J. *J. hum. Evol.* **19**, 699–736 (1990).
28. Stringer, C. B. & Andrews, P. *Science* **239**, 1263–1268 (1988).
29. Weidenreich, F. *Anthrop. pap. Am. Mus. Nat. Hist.* **43**, 205–290 (1951).
30. Murrill, R. I. *Z. Morph. Anthropol.* **66**, 176–187 (1975).
31. Aiello, L. C. & Dean, M. C. *Introduction to Human Evolutionary Anatomy* (Academic, New York, 1990).
32. Sartono, S. in *Paleoanthropology, Morphology and Paleoeology* (ed. Tuttle, R. H.) 327–360 (Mouton, The Hague, 1975).

ACKNOWLEDGEMENTS. T.L. thanks Jia Lanpo for encouragement, Hu Chengzhi for preparing the specimens and J. D. Clark, F. C. Howell and T. D. White for their hospitality during a recent visit to the Laboratory for Human Evolutionary Studies, University of California, Berkeley. D.A.E. thanks Jia Lanpo, Hu Chengzhi and Chen Zhenyu for their hospitality during recent visits to China. This research was supported by a grant from the L.S.B. Leakey Foundation.

Animal model of Gaucher's disease from targeted disruption of the mouse glucocerebrosidase gene

V. L. J. Tybulewicz*†, M. L. Tremblay†‡, M. E. LaMarca§, R. Willemssen||, B. K. Stubblefield§, S. Winfield§, B. Zablocka§, E. Sidransky§, B. M. Martin§, S. P. Huang‡, K. A. Mintzer*, H. Westphal‡, R. C. Mulligan‡ & E. I. Ginns§¶

* Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

‡ Laboratory of Mammalian Genes and Development, National Institute of Child Health and Human Development, Bethesda, Maryland 20892, USA

§ Section on Molecular Neurogenetics, Clinical Neuroscience Branch, National Institute of Mental Health, ADAMHA, Bethesda, Maryland 20892, USA

|| Department of Cell Biology and Genetics, Erasmus University, Rotterdam 3000 DR, The Netherlands

GAUCHER'S disease is the most prevalent lysosomal storage disorder in humans and results from an autosomally inherited deficiency of the enzyme glucocerebrosidase (β -D-glucosyl-N-acylsphingosine glucosylhydrolase)^{1–6}, which is responsible for degrading the sphingolipid glucocerebroside. An animal model for Gaucher's disease would be important for investigating its phenotypic diversity and pathogenesis and for evaluating therapeutic approaches. A naturally occurring canine model has been reported but not propagated⁷. Attempts to mimic the disease in animals by inhibiting glucocerebrosidase have been inadequate⁸. Here we generate an animal model for Gaucher's disease by creating a null allele in embryonic stem cells through gene targeting and using these genetically modified cells to establish a mouse strain carrying the mutation^{9,10}. Mice homozygous for this mutation have <4% of normal glucocerebrosidase activity, die within twenty-four hours of birth and store glucocerebroside in lysosomes of cells of the reticuloendothelial system.

To disrupt the murine glucocerebrosidase gene, a targeting plasmid was constructed containing a neomycin-resistance gene

Received 20 September 1991; accepted 16 March 1992.

1. Li, T., Wang, Z., Li, W., Feng, X., Hu, K. & Liu, W. *Jiangnan Kaogu* **39**, 1–14 (1991).
2. Black, D. *Palaeont. Sin. Ser. D* **7**, II, 1–144 (1931).
3. Weidenreich, F. *Paleont. Sin. Ser. D* **7** III, 1–162 (1936).
4. Weidenreich, F. *Palaeont. Sin. New Ser. D* **1**, 1–180 (1937).
5. Weidenreich, F. *Palaeont. Sin. Ser. D* **10**, 1–209 (1943).
6. Woo, J. *Sci. Sin.* **14**, 1032–1035 (1965).
7. Wu, R. & Dong, X. *Acta Anthropol. Sin.* **1**, 2–13 (1982).
8. Dong, X. in *Early Humankind in China* (eds Wu, R., Wu, X. & Zhang, S.) 9–23 (Science Press, Beijing, 1989).
9. Wu, X. *Sci. Sin.* **24**, 530–539 (1981).
10. Wu, R. *Acta Anthropol. Sin.* **7**, 97–101 (1988).
11. Woo, J. & Peng, R. *Palaeovert. et Palaeoanthropol.* **1**, 159–164 (1959).
12. Day, M. H., Leaky, M. D. & Magori, C. *Nature* **284**, 55–56 (1980).
13. Stringer, C., Howell, F. C. & Melentis, J. K. *J. archaeol. Sci.* **6**, 235–253 (1979).

† Present addresses: National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK (V.L.J.); Department of Biochemistry, McGill University, Montreal, Quebec H3G 1X6, Canada (M.L.T.).

¶ To whom correspondence should be addressed.

inserted into exons 9 and 10 (Fig. 1a), which encode part of the active site of the enzyme^{2,11}. A herpes simplex virus thymidine kinase gene flanks the construct, allowing the use of a positive-negative selection scheme¹². This construct was introduced into embryonic stem cells by electroporation and the cells were subjected to selection with the drugs G418 and gancyclovir¹². The correct gene targeting event in doubly resistant individual clones was identified by Southern blot analysis (Fig. 1b). Of 528 clones screened, 19 had undergone the homologous recombination event. As gancyclovir selection enriched the recombinants by 8.5-fold, the ratio of correctly targeted genes was one per 236 stably transformed cells.

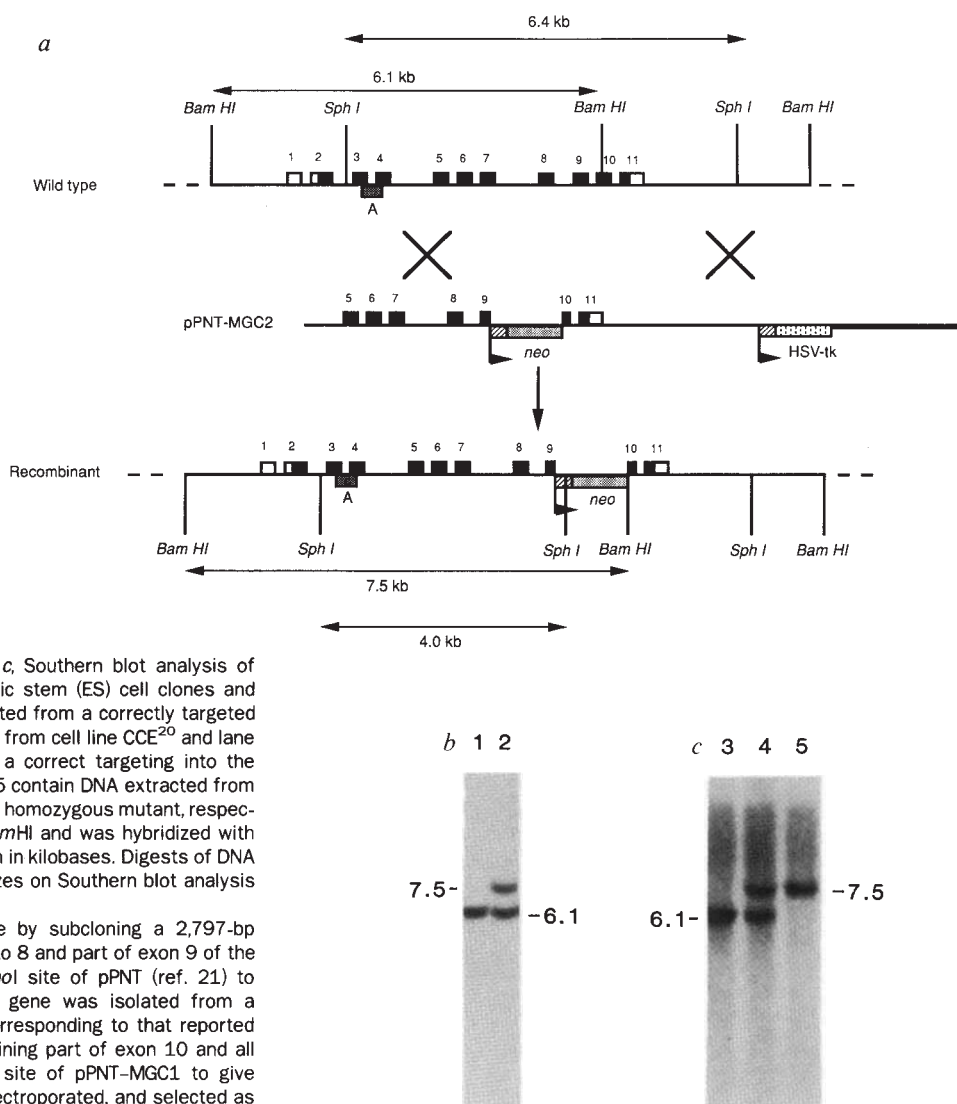
Cells from four embryonic stem cell clones containing the targeted mutation in the glucocerebrosidase gene of one allele were injected into blastocysts from C57BL/6 mice and transferred to FVB/N foster mice¹³. From these injections, male offspring with more than 30% chimaerism (as judged by coat colour) were test-bred against C57BL/6 females. One chimaeric male, derived from embryonic stem cell line 23-2A16, transmitted the glucocerebrosidase mutation to his progeny. Mice heterozygous for the disrupted glucocerebrosidase gene were mated and homozygous mutant progeny were identified by Southern blot analysis (Fig. 1c). As expected, the targeted mutation was transmitted in a mendelian fashion. Assay of glucocere-

brosidase activity from mouse tails demonstrated that in heterozygous mutants the mean activity was 44.4% (s.d. $\pm 2.2\%$) of that in normal littermates (Fig. 2a). The glucocerebrosidase activity in homozygous mutant mice was less than 4% of control, confirming that the targeted gene disruption had resulted in a null allele. Thin-layer chromatography of the neutral glycosphingolipids in tissues of normal and homozygous mutant mice demonstrated increased glucocerebroside levels only in the mutants (Fig. 2b).

Homozygous mutant mice are severely compromised at birth, are under weight and respire abnormally with rapidly progressing cyanosis, and their feeding and movement are decreased (Fig. 3). All homozygous mutant mice ($n > 100$) died within 24 h of birth. Using electron microscopy, macrophages with lysosomal lipid accumulation were found in the liver (Fig. 4a), bone marrow (Fig. 4b), spleen and brain of homozygous mutant mice ($n = 4$), but not in normal littermates (Fig. 4c; $n = 3$) or a heterozygous mutant. Twenty-five macrophages were examined in each tissue specimen. The overall architecture of the liver of the homozygous mutant mice appeared appropriate for the early neonatal stage of development, demonstrating extensive hepatic haematopoiesis. As in the human disease, hepatocytes were not affected and storage was restricted to the Kupffer cells in the liver sinusoids¹. Stored lipid in the macrophages of the

FIG. 1 a, Strategy for disruption of the murine glucocerebrosidase gene. The line labelled 'wild type' represents the normal genomic structure of the mouse glucocerebrosidase gene¹¹. Exons 1 to 11 are shown as boxes above the line, with filled boxes representing coding regions. The middle line labelled pPNT-MGC2 represents the targeting plasmid linearized at a unique *NotI* site. Neomycin-resistance (*neo*) and herpes simplex virus thymidine kinase (HSV-tk) genes are shown, both transcribed from the phosphoglycerate kinase-1 gene promoter (shaded boxes with arrows); thick line represents the plasmid backbone. The lower line labelled 'recombinant' shows the structure of the glucocerebrosidase gene after a correct targeting event. Parts of exons 9 and 10 and all of intron 9 are replaced with the *neo* gene. The shaded box labelled 'A' represents a 325-bp *HindIII*-*NarI* fragment used as probe to identify gene targeting events. Predicted sizes (in kilobases, kb) of *Bam*HI-digested and *Sph*I-digested fragments hybridizing to this probe are shown. b and c, Southern blot analysis of *Bam*HI-digested genomic DNA from embryonic stem (ES) cell clones and from progeny of a germ-line chimaera generated from a correctly targeted clone. In b, lane 1 contains normal ES cell DNA from cell line CCE²⁰ and lane 2 contains DNA of clone 23-2A16, showing a correct targeting into the glucocerebrosidase gene. In c, lanes 3, 4 and 5 contain DNA extracted from tails of mice that are normal, heterozygous and homozygous mutant, respectively. DNA in all lanes was digested with *Bam*HI and was hybridized with probe A (see a). Sizes of fragments are shown in kilobases. Digests of DNA with *Sph*I gave fragments of the expected sizes on Southern blot analysis (data not shown).

METHODS. Construct pPNT-MGC2 was made by subcloning a 2,797-bp *SspI*-*SalI* fragment containing all of exons 5 to 8 and part of exon 9 of the murine glucocerebrosidase gene into the *XhoI* site of pPNT (ref. 21) to create pPNT-MGC1. The glucocerebrosidase gene was isolated from a BALB/c DNA phage library and has a map corresponding to that reported in ref 11. A 3,205-bp *Bam*HI fragment containing part of exon 10 and all of exon 11 was subcloned into the *Bam*HI site of pPNT-MGC1 to give pPNT-MGC2. ES cell line CCE²⁰ was grown, electroporated, and selected as described²¹, then injected into C57BL/6 blastocysts¹³. DNA was prepared from ES clones and mouse tails and analysed by Southern blotting as before²¹.



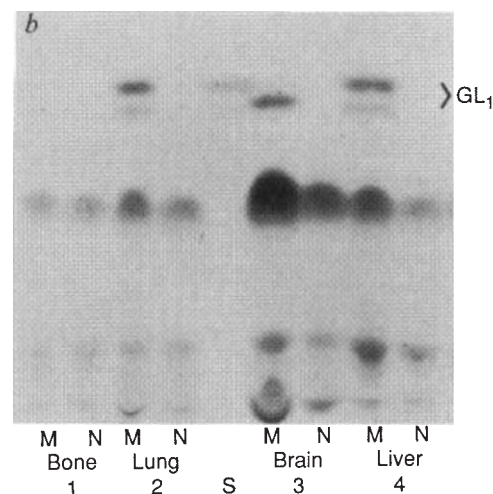
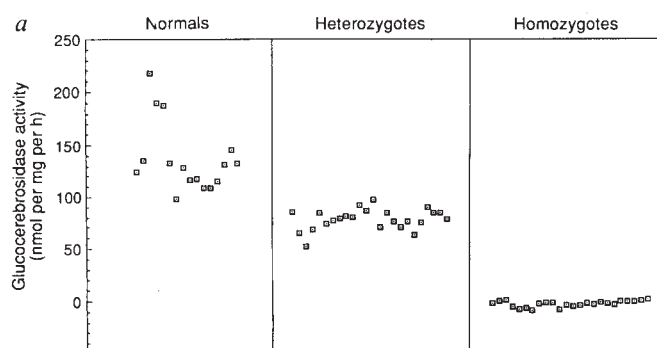


FIG. 2 *a*, Glucocerebrosidase activity in normal, heterozygous, and homozygous mutant mice. Enzyme activity was assayed in mouse tail extracts using 4-methylumbelliferyl- β -D-glucopyranoside as a substrate as described²². *b*, Thin-layer chromatography of neutral glycosphingolipids in tissues from normal and homozygous mutant mice. Bone marrow (lanes 1), lung (lanes 2), brain (lanes 3) and liver (lanes 4) from homozygous mutant (M) or normal (N) mice were extracted and analysed (see Methods). Results shown are from normal and homozygous mutant littermates and are representative of data from four homozygous mutant and normal pairs of littermates. The lane marked 'S' contains 4 μ g of standard glucocerebroside (GL₁). The position of glycosphingolipid GL₁, running as two bands on the chromatogram, is indicated. Glucocerebroside in brain differs from that in extraneural tissues, having predominantly, Glucocerebroside rather than C₂₂ fatty acid ceramide moieties²³. As a result of these differences in the fatty acid component, GL₁ resolves into two components.

METHODS. Tissues from normal and homozygous mutant mice were ex-

tracted with 9.5 ml (v/v) of chloroform/methanol (2:1 v/v) with sonication for 10 min at room temperature followed by addition of 0.2 volumes water. After vortexing, samples were centrifuged for 2 min. A 100- μ l aliquot of the lower phase containing the neutral glycosphingolipids from each tissue was evaporated to dryness, resuspended in 50 μ l chloroform/methanol (2:1 v/v) and applied to precoated silica gel-60 high-performance thin-layer chromatography plates (Merck) that had been activated at 100 °C for 1 h²⁴. The solvent mixture used for development was chloroform-methanol-water, 65:25:4 (v/v/v). After development for 30 min, the plates were air-dried, and the glycolipids visualized with orcinol-ferric chloride (Bial's reagent)²⁴.

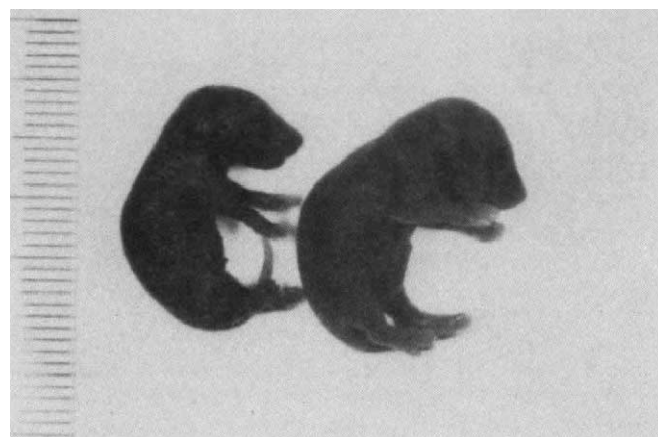
homozygous mutant mice has the same physical appearance as the tubular lysosomal deposits of glucocerebroside in cells of human patients with Gaucher's disease (Fig. 4d)^{14,15}.

Although all homozygous mutant mice succumb within the first day of life, no single factor causing death has been identified. Glucocerebroside is estimated by thin-layer chromatography to be 5–10 μ g per mg wet weight of liver, lung and brain tissue of mutant homozygous mice at birth (Fig. 2b). Less striking amounts of glucocerebroside are seen in bone marrow (Fig. 2b) and plasma. Electron microscopy has demonstrated that the storage in lysosomes of macrophages is present but not severe. Thus, despite the consistent finding of increased tissue glucocerebroside and storage within macrophages in the liver and spleen, there does not appear to be sufficient hepatosplenomegaly to explain the rapid deterioration in these newborn mutant mice. Soon after birth the homozygous mutant mice are always separated from healthy littermates by their mothers and neglected, so both the underlying pathology and the maternal perception of abnormality probably contribute to the early demise of the newborn mutant mice.

The irregular respiration, poor feeding and decreased movement of the mutant mice are consistent with the nervous system dysfunction responsible for the fulminant clinical course seen in the more severely affected subset of type-2 (acute neuronopathic) infant Gaucher's patients^{1,16,17}. Our results may therefore indicate that null mutations similar to the one created here could be responsible for the early mortality in this subset of infants with type-2 Gaucher's disease who die of neurological disease shortly after birth, so we would predict that individuals homozygous for these null mutations would not survive beyond early infancy. Mutant mice carrying less deleterious mutations could therefore be models for the more frequently observed milder presentations of Gaucher's disease.

Mice negative for the enzyme hypoxanthine ribosyl phosphoryl transferase have been generated by gene targeting, but these were asymptomatic and were not adequate as a model for Lesch-Nyhan disease^{18,19}. The mice described here may thus be the first example of a mouse model of a human genetic disease generated by gene targeting. These mice should be useful for studying the pathogenesis and the phenotypic diversity seen

FIG. 3 Photograph of a homozygous mutant mouse (left) and a normal littermate (right) ~6 h after birth. In contrast to the normal mouse, the mutant mouse is cyanotic, akinetic, has not fed and has abnormal skin (scale graduated in millimetres).



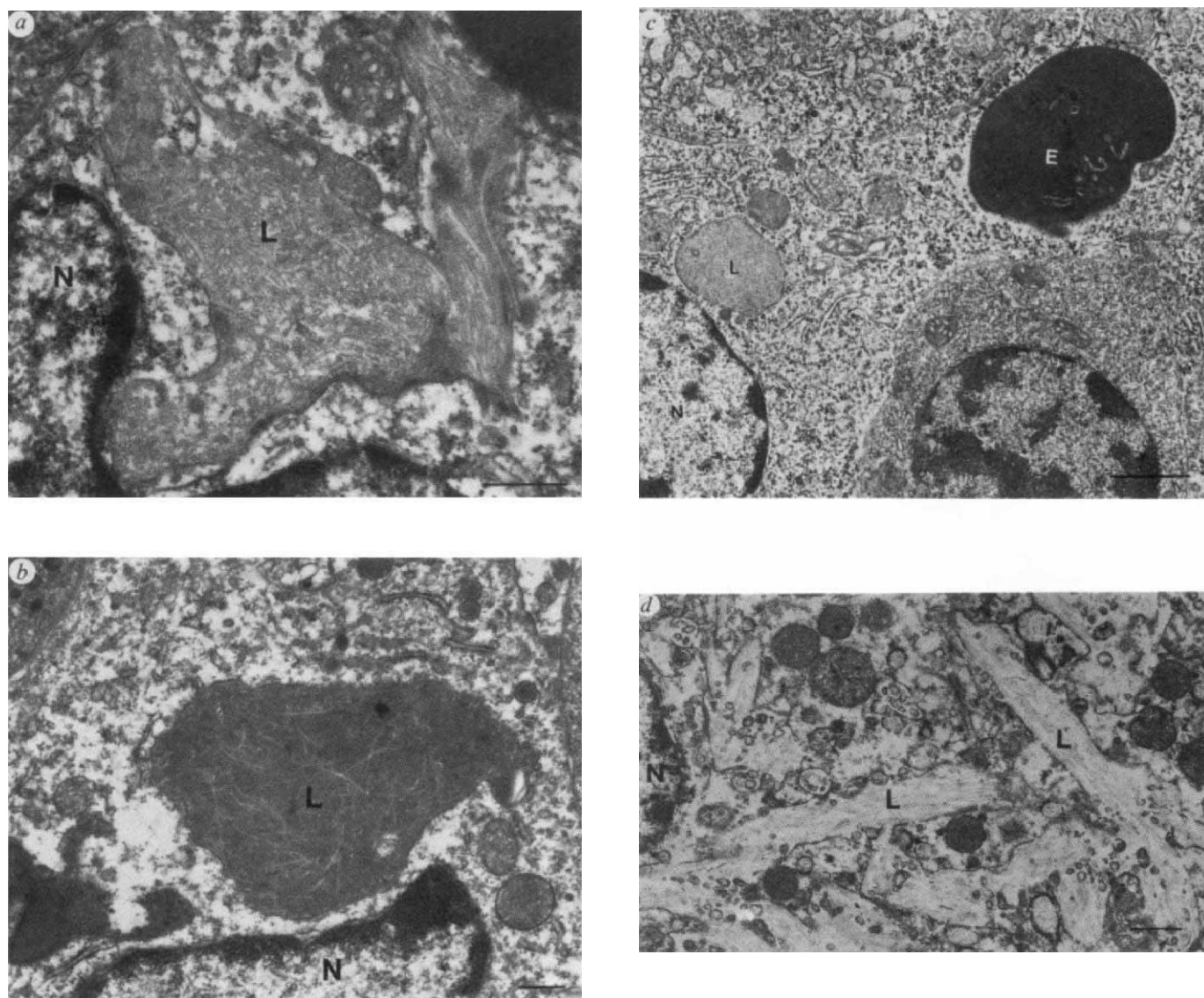


FIG. 4 Electron micrographs showing the characteristic morphology of lipid storage material in lysosomes of macrophages in the liver (a) and bone marrow (b) of homozygous mutant mice with glucocerebrosidase deficiency. In contrast to the more rounded lysosome morphology in normal newborn mice (c), the tubular and elongated shape of stored lipid gives the lysosomes in homozygous mutant mice a more irregular shape. Stored glucocerebroside

in the lysosome of a macrophage in the spleen of an affected human fetus is shown in d (kindly provided by V. D. Vuzevski). L, lysosome; E, phagocytosed erythrocyte; N, nucleus. Scale bars, 0.5 μ m. METHODS. Tissues were fixed in phosphate buffer, pH 7.3, containing 1% acrolein and 0.4% glutaraldehyde and postfixed according to ref. 25. Standard procedures were used for Epon embedding.

in Gaucher's disease and will provide a model in which enzyme replacement, cellular transplantation and gene transfer therapies can be evaluated. □

Received 27 February; accepted 30 April 1992.

1. Barranger, J. A. & Ginns, E. I. in *The Metabolic Basis of Inherited Disease* (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 1677-1698 (McGraw-Hill, New York, 1989).
2. Martin, B. M., Sidransky, E. & Ginns, E. I. *Adv. Pediatr.* **36**, 277-306 (1989).
3. Tsuji, S. *et al. New Engl. J. Med.* **316**, 570-575 (1987).
4. Tsuji, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 2349-2352 (1988).
5. Beutler, E., Gelbart, T., Kuhl, W., Sorge, J. & West, C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 10544-10547 (1991).
6. Grabowski, G. A., Gatt, S. & Horowitz, M. *Crit. Rev. Biochem. molec. Biol.* **25**, 385-414 (1991).
7. Hartley, W. J. & Farrow, R. R. H. in *Handbook: Animal Models of Human Disease. Fasc II*, (ed. Capen, C. C. *et al.*) Model No. 241, Registry of Comparative Pathology (Armed Forces Institute of Pathology, Washington, DC, 1982).
8. Kanfer, J. N., Stephens, M. C., Singh, H. & Legler, G. *Prog. clin. biol. Res.* **95**, 627-644 (1982).
9. Robertson, E. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 71-112 (IRL, Oxford, 1987).
10. Capecchi, M. R. *Trends Genet.* **5**, 70-76 (1989).
11. O'Neill, R., Tokoro, T., Kozak, C. A. & Brady, R. O. *Proc. Natn. Acad. Sci. U.S.A.* **86**, 5049-5053 (1989).

12. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. *Nature* **336**, 348-352 (1988).
13. Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 113-151 (IRL, Oxford, 1987).
14. Willemsen, R. *et al. Ultrastr. Path.* **12**, 471-478 (1988).
15. Lee, R. E. *Prog. clin. biol. Res.* **95**, 177-218 (1982).
16. Fredrickson, D. S. & Sloan, H. R. in *The Metabolic Basis of Inherited Disease* (eds Stanbury, J. B., Wyngaarden, J. B. & Fredrickson, D. S.) 731-759 (McGraw-Hill, New York, 1978).
17. Sun, C. C., Panny, S., Combs, J. & Gutberlett, R. *Path. Res. Pract.* **179**, 101-104 (1984).
18. Hooper, M. L., Hardy, K., Handyside, A., Hunter, S. & Monk, M. *Nature* **326**, 292-294 (1987).
19. Kuehn, M. R., Bradley, A., Robertson, E. J. & Evans, M. J. *Nature* **326**, 295-298 (1987).
20. Robertson, E., Bradley, A., Kuehn, M. & Evans, M. *Nature* **323**, 445-448 (1986).
21. Tybulewicz, V. L. J., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. *Cell* **65**, 1153-1163 (1991).
22. Beutler, E. & Kuhl, W. L. *Lancet* **1**, 612-613 (1970).
23. Svennerholm, L., Hakansson, G., Mansson, J. E. & Nilsson, O. in *Gaucher's Disease: A Century of Delineation and Research* (eds Desnick, R. J. *et al.*) 231-252 (Liss, New York, 1982).
24. Kundu, S. K. in *Meth. Enzym.* (ed. Lowenstein, J. M.) 185-204 (Academic, New York, 1981).
25. Debruyjn, W. C. & Denbreejen, P. *Histochem. J.* **8**, 121-142 (1976).

ACKNOWLEDGEMENTS. We thank H. Stewart, A. Liebelt, T. O'Neill, R. Lee and M. Amver for, histopathological consultations, A. Reuser, W. Eliason, G. Bingham, K. Mahon, J. LeGross and B. Martin for help and discussion, and E. Alzona for typing and editorial assistance. This work was supported in part by the National Gaucher Foundation, by a postdoctoral fellowship (V.L.J.T.) from the Irvington Institute for Medical Research, a grant from the NIH (R.C.M.), and by a postdoctoral fellowship from the National Cancer Institute of Canada (M.L.T.).

Mechanical performance of scallop adductor muscle during swimming

Richard L. Marsh, John M. Olson & Sharon K. Guzik

Northeastern University, Department of Biology, 360 Huntington Avenue, Boston, Massachusetts 02115, USA

MECHANICAL performance of skeletal muscle has long been the subject of intense interest¹, but the details of *in vivo* performance of individual skeletal muscles during normal locomotion remain largely unknown. Performance *in vitro* has been described with considerable precision under simplified loading conditions². The force production and shortening velocity of most muscles, however, probably change continuously during natural movements. Therefore, modelling *in vivo* performance on the basis of *in vitro* contractile properties³ is subject to large degrees of uncertainty. Designing *in vitro* experiments that effectively examine the limits of mechanical performance requires increasing knowledge of precisely how muscles are used during normal movements. We report here measurements of the mechanical performance of the adductor muscle in scallops during jet-propulsion swimming. Swimming in scallops is powered solely by the striated portion of the single adductor muscle. Exploiting this simple locomotor morphology with simultaneous high-resolution measurements of pressure and flow rate, we have recorded nearly instantaneous measurements of the performance of a single skeletal muscle during normal locomotion.

We studied swimming in two species of comparably sized small scallops, *Argopecten irradians* and *Chlamys hastata*. Propulsive power in scallops is formed by the rapid closing (adduction) of the two valves, caused by the nearly simultaneous contraction of the entire striated adductor muscle⁴⁻⁷ (R.L.M. and J.M.O., unpublished results). Water escaping through jets near the hinge propels the animal ventral (gape) side first through the water. The brief pressure pulses during swimming

were measured with a small catheter-tip pressure sensor designed for intravascular use in mammals. Accurate estimates of flow rate (change in volume per time) were made possible by the simple geometry of scallops. Volume change was estimated by measuring the distance between the valves with sonomicrometer transducers and calculating the change in volume.

Both the pressure produced by the adductor muscle and its shortening velocity varied continuously during swimming (Fig. 1). Initial adduction of the valves was rapid because the mantle is unsealed (Fig. 1a). The valves then decelerated rapidly as the mantle came together and pressure rose rapidly. After reaching a peak of about 3–4 kPa, pressure fell because of relaxation of the adductor muscle, which is activated in a twitch-like manner during swimming (ref. 7 and R.L.M., S.K.G. and J.M.O., unpublished results). Flow rate and pressure were positively correlated after the peak pressure, but not before. Pressure fell below ambient during opening as water was drawn into the shell. Abduction was terminated abruptly, presumably limited by the preset length of the smooth portion of the adductor muscle⁶. Another adduction began after a brief pause, 38 ± 2.7 (s.e.m.) and 23 ± 2.7 ms for *Argopecten* and *Chlamys*, respectively.

Power calculated from pressure and flow rate (Fig. 1c) reached a very sharp peak soon after the formation of the jet. Peak power output was quite high given the low temperature at which these muscles were operating ($\sim 10^\circ\text{C}$). Power averaged over the entire cycle was 23 and 31% of the peak value in *Argopecten* and *Chlamys*, respectively (Fig. 2). During opening, positive power was also calculated because the sign of pressure and flow were both reversed. This pressure-volume work represents the release of energy stored in the elastic hinge ligament⁸. Between the two species power output varies directly with frequency as stroke work is quite similar (Fig. 3).

The highest performance during swimming occurred during the first 'clap' of the valves (Figs 2 and 3). Stroke work dropped by an average of 10 and 15% from the first to the second clap in *Argopecten* and *Chlamys*, respectively. Performance continued to decline at a lower rate with subsequent adductions. Decreased performance between the first and second claps was partly due to a difference in the dynamics of valve adduction

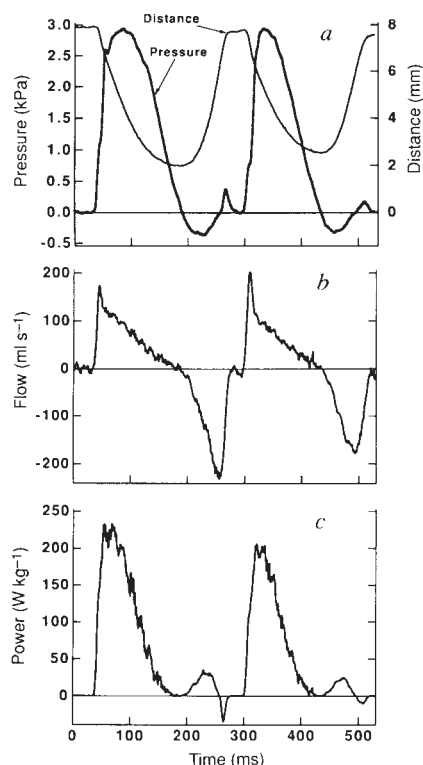


FIG. 1 Example of data collected during the first two valve claps in *Chlamys hastata*. a, Pressure in the mantle cavity and distance between the valve margins at the level of the adductor muscle (fully closed = 0 mm); b, flow rate of water out of the mantle cavity; c, mechanical power calculated as the product of pressure (Pa) and flow ($\text{m}^3 \text{s}^{-1}$) divided by the mass (kg) of the striated adductor muscle. All data were collected with a sample period of 0.833 ms using a MacAdiosII 12 bit A-D converter and Superscope software from G.W. Instruments running in a Macintosh II computer. Pressure was measured with a Millar model SPR-407 catheter tip pressure sensor. The sensor was inserted through a 1.3-mm internal diameter guide cannula glued into the upper (left) valve. The site of pressure measurement was about half the distance from the hinge to the ventral margin of the valve. Distance was measured using 1 or 1.5 mm diameter piezoelectric transducers connected to a Triton Technology Model 120 sonomicrometer. The sonomicrometer was recalibrated for the speed of sound in sea water at 10°C . The transducers were located adjacent to the anterior side of the adductor muscle. The distance trace has been shifted to the left by 5 ms, which corrects for the linear phase shift introduced by the active filters in the sonomicrometer (R.L.M., unpublished results). Volume was calculated from the distance trace after correcting for the distribution of projected shell area in relation to the centre of rotation at the hinge and the position of the length transducers. This correction factor was calculated by performing a strip analysis in 2-mm increments using digitized images. High-speed cine film was used to verify that the mantle does not bulge significantly during the high-pressure phase of closing, and thus the wedge of volume enclosed by the projected margins of the shell accurately reflects the volume enclosed by the mantle.

(Fig. 1, R.L.M. *et al.*, unpublished results). Subsequent declines were almost certainly due to fatigue of the adductor muscle, which is fuelled by anaerobic pathways⁹.

Have we quantified most of the work done during swimming in scallops by quantifying the pressure-volume work? Errors due to the kinetic energy of the water near the site of pressure measurements were minimized by keeping the pressure transducer well away from the jets¹⁰. Two power requirements in addition to the pressure-volume work also need to be considered. First, power is required to accelerate the apparent mass

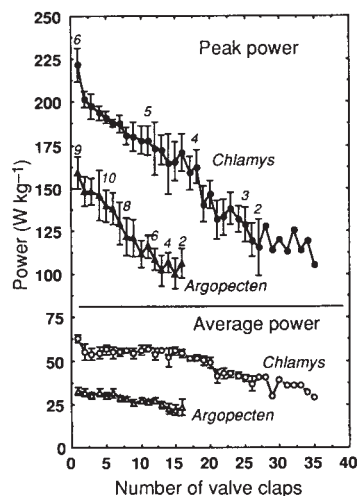


FIG. 2 Mean power produced by the striated adductor muscle as a function of the number of valve claps during a swim. The error bars represent ± 1 s.e.m. The sample size varies with pulse number because animals stopped swimming after variable numbers of claps and the first pulses on one swim in *Argopecten* were not usable because of noise. The numbers above the peak power curves indicate the sample size at several points. *Argopecten* swam in response to placing a predatory snail (*Nucella*) near the hinge; *Chlamys* swam after being touched with a tube foot from a starfish (*Pycnopodia*). The data shown are for first swims on the day of measurement. Average power was determined by integrating the power trace for each clap with respect to time and dividing by the period. *Argopecten* were collected off the southern coast of Cape Cod, by the Department of Marine Resources at Marine Biological Laboratory, Woods Hole. *Chlamys* were collected near the University of Washington, Friday Harbor Marine Laboratory on Puget Sound. The mass of the animals used in these measurements was 27.2 ± 4.0 and 17.0 ± 1.7 for *Argopecten* and *Chlamys*, respectively. The valve dimensions for *Argopecten* and *Chlamys* were, respectively: dorsal-ventral, 50.1 ± 2.6 and 52.0 ± 0.7 mm; anterior-posterior, 50.8 ± 2.8 and 44.9 ± 0.4 mm. The striated muscle weighed 1.95 ± 0.31 and 1.55 ± 0.06 g in *Argopecten* and *Chlamys*, respectively.

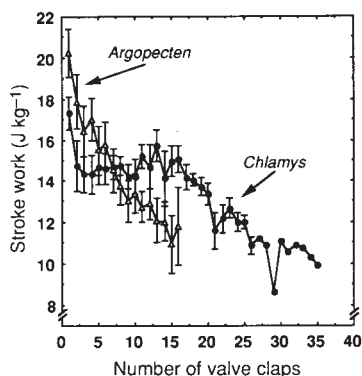


FIG. 3 Stroke work of the striated adductor muscle as a function of the number of valve claps during a swim. The error bars represent ± 1 s.e.m. Sample sizes are as in Fig. 1. The stroke work was calculated by integrating the power trace for each clap.

of the valves, including the added-mass effect¹¹. All of this power is expended during a brief period (< 15 ms) very early in adduction before the mantle seals (Fig. 1). Subsequently, the kinetic energy of the valves is returned as the valves decelerate during jetting, a reaction that should contribute to pressurizing the mantle cavity. But, using maximal estimates for the amount of power involved in acceleration and deceleration resulted in only a slight change in the shape of the power curve early in adduction (Fig. 4). Second, work is required to compress the hinge. As indicated above, a portion of this work (a mean of 0.57 J kg^{-1}

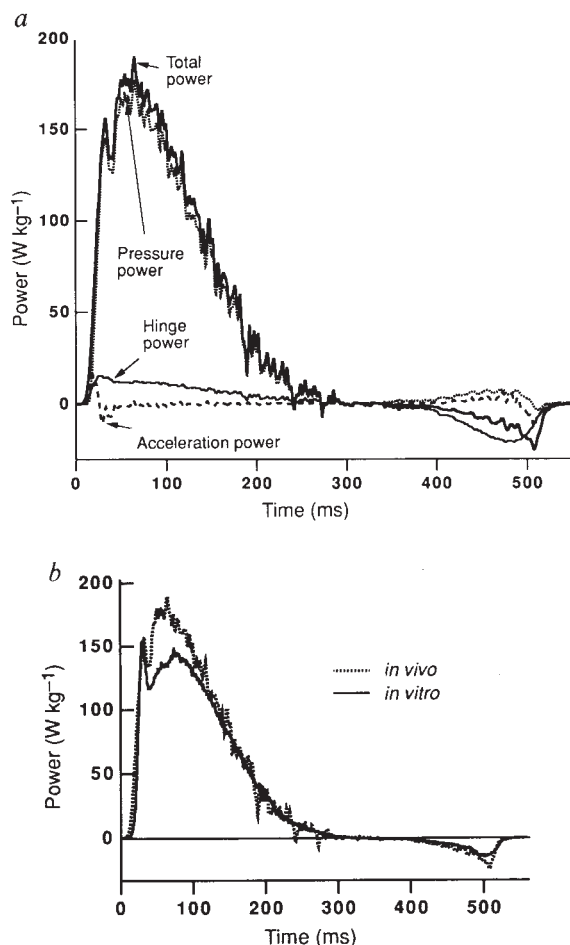


FIG. 4 *a*, Power produced by the adductor muscle during the first cycle of a swim in *Argopecten* at 10°C . The curve labelled 'Pressure Power' was calculated from pressure and flow rate using the methods described in Fig. 1. 'Hinge power' is the power stored and released by the hinge, calculated from the velocity of shortening and the force required to compress the hinge (R.L.M. and J.M.O., unpublished results). 'Acceleration power' is the power needed to accelerate the valves. This power includes a calculation of the added mass effect assuming that the valves act like flat plates¹². The added mass calculation results in an equivalent valve mass of about 12 times the actual mass of the valves. This calculation may well be an overestimate of the acceleration power as the pressure sensor, located on the inner surface of the left valve, was probably influenced by the acceleration reaction. 'Total power' is the sum of the three components of power. The power curve during adduction is dominated by the very large contribution of pressure power. During opening the pressure power and the acceleration power are satisfied by power released from the hinge. *b*, Comparison of the total power calculated *in vivo* with the power measured *in vitro* using an isolated strip of adductor muscle (R.L.M. and J.M.O., manuscript in preparation). The isolated muscle was put through a length cycle that mimics that occurring *in vivo* and was stimulated at a phase determined by *in vivo* electromyographic recording. Qualitatively the two measurements agree quite precisely, including the power required to reextend the muscle during abduction. Quantitatively, the *in vitro* power averaged $\sim 20\%$ lower than that estimated *in vivo*, but the difference was not statistically different (R.L.M. and J.M.O., manuscript in preparation).

adductor muscle in *Argopecten*) was accounted for as the pressure-volume work required to refill the mantle cavity during abduction. The total work needed to compress the hinge in *Argopecten* is about 1–1.5 J kg⁻¹ muscle (R.L.M. and J.M.O., unpublished results), and thus our estimates of stroke work based on pressure and volume are low by 0.4–0.9 J kg⁻¹ or only about 2–5%. Redistributing the extra work required to compress the hinge during shortening also resulted in very little change in the curve (Fig. 4). Therefore, our measurements of performance during jet propulsion in scallops probably reflect very closely and almost instantaneously (time resolution of <5 ms out of a total cycle time of 250–600 ms) the actual performance of the adductor muscle during swimming.

We know of no comparable measures of performance with this time resolution for single skeletal muscles. Previous measurements of swimming performance in scallops are of much lower resolution and also greatly underestimate the pressure produced¹². Maximal performance has been measured precisely

for human movements many times^{1,13}, but the recruited mass of muscle is difficult to determine accurately. Also, deducing the performance of a single muscle during complex movements in man or other vertebrates is fraught with uncertainty¹⁴. Performance of individual insect muscles that provide the energy for jumping has been estimated¹⁵, but the calculations involve several indirect measurements and assumptions.

The data on *in vivo* performance collected here also provide new avenues for evaluating the *in vitro* performance of skeletal muscles. The accurate length cycles measured with the sonomicrometer have allowed us to design *in vitro* contractile experiments that mimic the *in vivo* contractions (Fig. 4b). These muscles operate on their isotonic force-velocity curve for only a brief period during shortening. Further correlating the details of *in vivo* lengthening and shortening cycles with the intrinsic properties of individual muscles will be important for constructing and evaluating models of the performance limits of striated muscle. □

Received 27 February; accepted 6 April 1992.

- Hill, A. V. *J. Physiol., Lond.* **53**, lxxxviii–xc (1920).
- Wolledge, R. C., Curtin, N. A. & Homsher, E. *Energetic Aspects of Muscle Contraction* (Academic, New York, 1985).
- VanLeeuwen, J. L. *J. theoret. Biol.* **149**, 229–256 (1991).
- Yonge, C. M. *Mem. Mus. R. Hist. Nat. Belg.* **3**, 77–100 (1936).
- Lowy, J. *J. Physiol., Lond.* **124**, 100–105 (1954).
- Wilkins, L. A. *Proc. R. Soc. B* **211**, 341–372 (1981).
- Mellon, D. Jr *Z. vergl. Physiol.* **62**, 318–336 (1969).
- Trueman, E. R. *J. exp. Biol.* **30**, 453–467 (1953).
- Gade, G., Weeda, E. & Gabbott, P. A. *J. comp. Physiol.* **124**, 121–127 (1978).
- Wang, C. Y. & Sonnenblick, E. H. *J. Biomech.* **12**, 9–12 (1979).

- Daniel, T. L. *Am. Zool.* **24**, 121–134 (1984).
- Moore, J. D. & Trueman, E. R. *J. exp. mar. Biol. Ecol.* **6**, 179–185 (1971).
- Davies, C. T. M. & Sandstrom, E. R. *Eur. J. appl. Physiol.* **58**, 838–844 (1989).
- Edgerton, V. R., Roy, R. R., Gregor, R. J. & Rugg, S. in *Human Muscle Power* (eds Jones, N. L., McCartney, N. & McComas, A. L.) 43–64 (Human Kinetics, Champaign, Illinois, 1986).
- Bennet-Clark, H. C. *J. exp. Biol.* **63**, 53–83 (1975).

ACKNOWLEDGEMENTS. We thank the staff of the Friday Harbor Marine Laboratory for facilitating the work on *Chamlys*, S. Norton, E. Kozloff, D. Janies and C. Staude for collecting *Chamlys* and for crucial apparatus, K. Lee, J. Dumas and K. Paull for *Nucella* and T. Daniel and S. Vogel for suggestions regarding the calculation of added mass effects. This study was supported by a grant from the NIH to R.L.M. and a Biomedical research Support Grant from the NIH to Northeastern University.

Function of maternal cytokeratin in *Xenopus* development

N. Torpey, C. C. Wylie & J. Heasman*

The Wellcome/Cancer Research Campaign Institute of Cancer and Developmental Biology, Tennis Court Road, Cambridge CB2 1QR, and Department of Zoology, Downing Street, Cambridge CB2 3ED, UK

INTERMEDIATE filaments are ubiquitous in eukaryotic cells, but their functions are poorly understood. The *Xenopus* oocyte contains both messenger RNA and protein products of cytokeratin and vimentin genes^{1–5} in non-overlapping arrays. The cytokeratin filaments contain dimers of the type I (acidic) subunit XLK3a(19), and the type II (basic) subunit XCK1(8), polymerized to form a cortical network. These are homologues of the human simple epithelial keratins 19 and 8, respectively. After the first few cell cycles following fertilization these filaments become restricted to the superficial cells of the blastula. We have depleted the oocyte's store of the type II cytokeratin mRNA by injecting antisense oligodeoxynucleotides (oligos) and studied the effect on embryonic development. As zygotic transcription does not commence until the late blastula stage⁶, there are at least 9 hours in which to see the effect of loss of function of this mRNA. We report here that the cytokeratin filaments become depleted in the cortical cells of the embryo. As a result, there is a loss of the 'compacted' epithelial surface of the blastula, an inability to close a wounded surface and defective gastrulation.

Six oligos complementary to different regions of the published sequence for the central rod-domain of the type II cytokeratin XCK1(8) were injected into full-grown (stage 6) ovarian oocytes (Fig. 1a). Only one oligo (CK5) reduced the level of XCK1(8) mRNA significantly and in a dose-dependent manner. As CK5 had to be injected in high doses (20 ng per oocyte) to obtain a greater than 90% depletion, a methoxyethylphosphoramidate-modified form of CK5 (CK5-GC) was synthesized⁷ (Integrated

TABLE 1 Number of embryos developing from oligo-injected oocytes and scored for the abnormal phenotype described in Fig. 3

(a) Blastula stage				
Experiments 1 and 2	0.2 ng CK5	1 ng CK5	1 ng AMD7	uninjected
Decompacted at the				
blastula stage	12/32	26/30	0/22	0/45
Normal at the				
blastula stage	17/32	4/30	15/22	31/45
(b) Gastrula stage				
Experiment 3	3 ng CK5	3 ng CK5S	3 ng CK5+ 3 ng CK5S	
Exogastrulae	20/22	3/22	7/30	
Normal gastrulae	2/22	19/22	23/30	

a, Numbers of embryos recovered in two experiments after injection of oocytes with modified oligos, and scored for the phenotype of 'cobblestone' appearance at the blastula stage compared to the total number of embryos at that stage. In both uninjected and experimental batches some embryos had non-specific abnormalities. In both experiments the degree of mRNA depletion was monitored by Northern analysis of eggs and embryos as described in Fig. 1b. CK5-GC (1.5 ng) depleted over 95% of the XCK1(8) mRNA (data not shown). b, Numbers of embryos in experiment 3, injected as oocytes with modified oligos, devitellined and scored for exogastrulation. Northern analysis showed that 3 ng CK5-GC depleted XCK1(8) mRNA almost completely, whereas the sense oligo and sense + antisense had no effect (data not shown).

DNA Technology). Figure 1b shows that 1.5 ng CK5-GC caused almost complete depletion of XCK1(8) mRNA. This knock-out was specific as irrelevant modified oligos did not have this effect. Furthermore, the amount of the mRNA for the type I keratin partner of XCK1(8), the acidic keratin XLK3a(19), was unaffected by CK(5)-GC injection.

Next we studied whether depletion of XCK1(8) mRNA caused an alteration in the amount of XCK1(8) protein synthesized in the oocyte. We know from previous studies with specific antibodies that XCK1(8) protein is the most abundant triton-insoluble protein in *Xenopus* oocytes and early embryos⁵. It is

* To whom correspondence should be addressed.

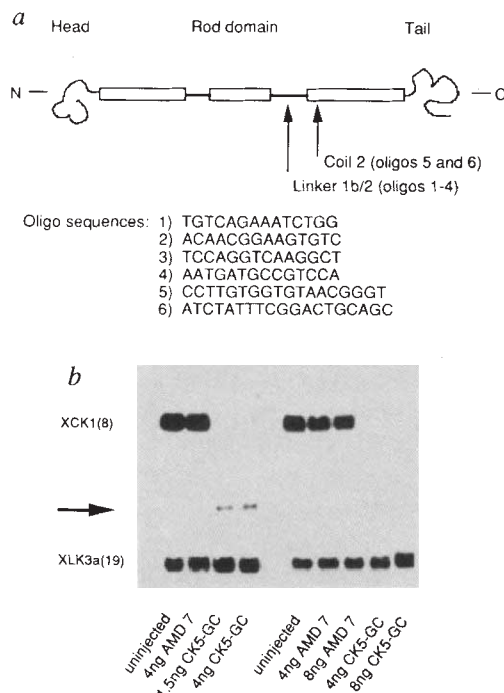


FIG. 1 **a**, A diagram of the XCK1(8) protein, indicating the regions of the cognate mRNA to which the antisense oligos should hybridize. Oligos 1-4 are all in the linker 1b/2 domain, 5 and 6 are in coil 2. Only oligo 5 (CK5) was effective in depleting the XCK1(8) mRNA. **b**, Northern blots of two experiments to test the degree of depletion of XCK1(8) mRNA in oocytes after injection of various oligos. AMD 7 is an antisense oligo effective in depleting mRNA for an unrelated protein, MyoD. 1.5 ng of CK5-GC (lane 3) is sufficient for almost complete depletion of XCK1(8) mRNA, whereas the irrelevant oligo and CK5S-GC (not shown) has no effect. Reprobing the blot with a probe for the type 1 keratin partner of XCK1(8) shows that CK5-GC does not deplete the mRNA of this type 1 cytokeratin XLK3a(19). Nor is vimentin mRNA depleted (data not shown). Notice the appearance of a new band when XCK1(8) is depleted (arrow). This was shown using fragment-specific RNase protection probes to be the stable 5' fragment of the cleaved mRNA (data not shown).

METHODS. Oocytes were manually defolliculated, injected with oligos and incubated overnight. RNA was extracted¹² and separated on formaldehyde-MOPS-agarose gels¹³. Two oocyte equivalents were loaded per lane. The RNA was capillary blotted onto GeneScreen membrane (NEN), and the filters were hybridized at 65 °C for 16 h using a random primed probe for XCK1(8) mRNA. After washing and exposing to film, filters were stripped and rehybridized with a probe for XLK3a(19). After washing they were exposed to film for 24 h and the autoradiogram was photographed.

clearly seen in an autoradiogram of ³⁵S methionine-labelled, newly synthesized proteins (Fig. 2a). By contrast, after CK5 injection the synthesis of XCK1(8) protein was significantly reduced (Fig. 2b). Furthermore, CK5-injected oocytes cultured for 60 h show significant reduction in the number and length of their cytokeratin filaments compared with controls, as shown by whole-mount immunocytochemistry (Fig. 2c, d).

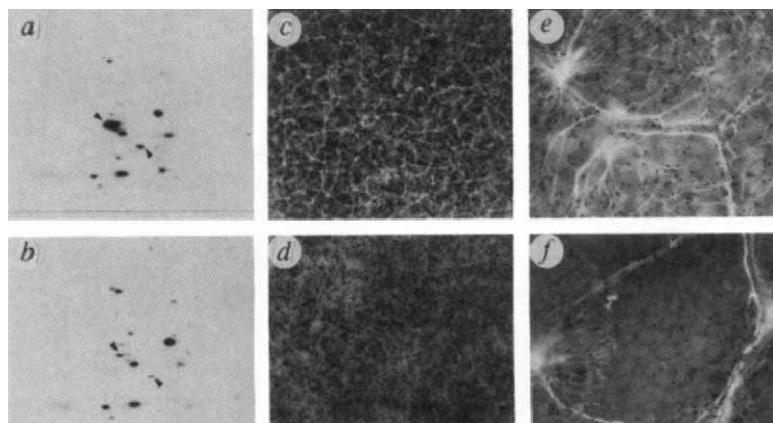
We next investigated whether cytokeratin mRNA-depleted oocytes would develop normally if fertilized, and whether there was a concomitant loss of cytokeratin protein in the early embryo. To fertilize CK5-GC-injected oocytes we used the host-transfer technique described previously⁸. As controls, we injected several oligos complementary to other known mRNAs as well as the sense version of CK5-GC (CK5 S-GC). Embryos injected as oocytes with 1-3 ng CK5-GC developed normally to the mid-blastula stage, even though RNase protection analysis showed at least 90% depletion of the maternal mRNA (data not shown). When wholemounts of embryos fixed at this stage were

stained with an anti-XCK1(8) antibody, they had fewer cytokeratin filaments in their superficial cells compared to controls. The remaining filaments were often shorter while still maintaining their association with the cell membrane (Fig. 2e, f).

At the mid-blastula stage, the superficial cells of the control embryos flattened over the embryos' surfaces, a process known as compaction (Fig. 3a). By contrast, those of the CK5-GC-injected embryos remained raised up, giving a 'cobblestone' appearance. Embryos with this phenotype began to gastrulate but the progress of gastrulation was delayed, and cells which normally invaginate failed to do so, spreading instead over the animal hemisphere inside the vitelline membrane. When their vitelline membranes were removed at the blastula stage, the embryos partially exogastrulated (Fig. 3c). This effect was not seen when irrelevant or CK5 S-GC oligos were used (Fig. 3d) and also was partially rescued when sense and antisense CK5-GC were injected together (Fig. 3e). The phenotype associated with depletion of XCK1(8) mRNA was highly reproducible

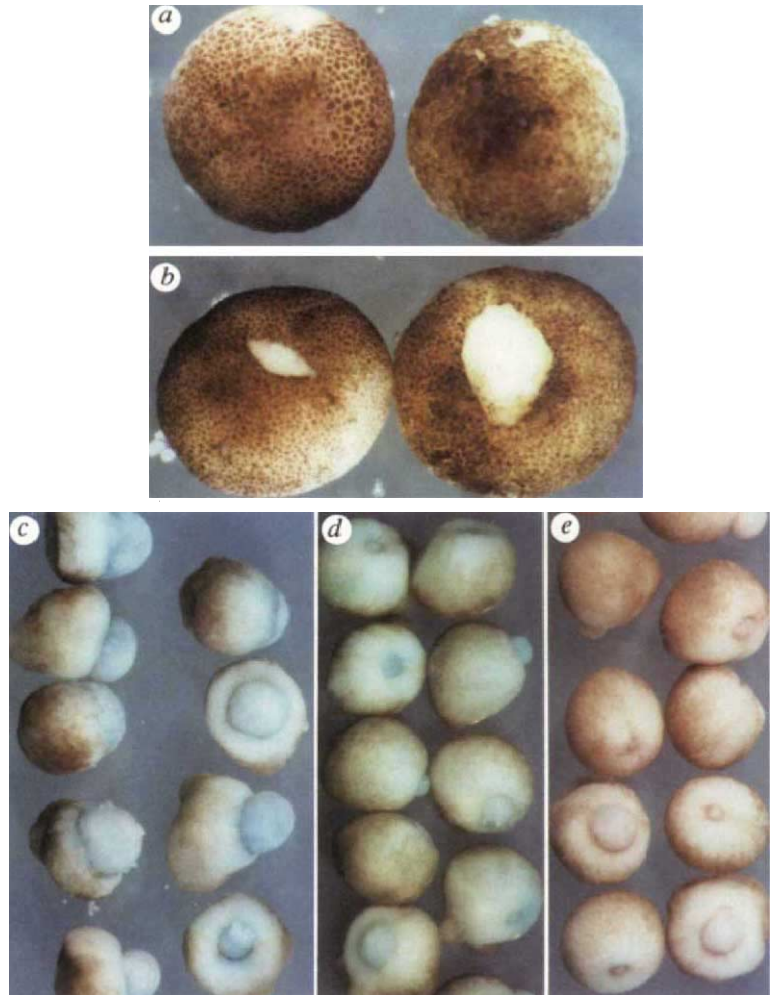
FIG. 2 The depletion of XCK1(8) mRNA in oocytes inhibits XCK1(8) protein synthesis and reduces protein filaments in oocytes and embryos. **a**, **b**, Oocytes injected with 18 ng CK6 or 18 ng CK5 oligos (both unmodified) were labelled with ³⁵S-methionine⁴ and triton extracts were analysed by two-dimensional electrophoresis and fluorography. Synthesis of XCK1(8) (*M*_r 57 K) is abundant in the control, but is much reduced in the XCK1(8) mRNA-depleted oocytes (**a**, **b**, upper arrowhead). Most other newly synthesized proteins are unaffected, although four smaller proteins are also reduced in amount (for example, lower arrowhead **b**). These may represent other keratins or keratin-related proteins made unstable by the loss of XCK1(8), as is known to be the case in other systems¹⁴⁻¹⁸. **c**, **d**, Whole-mount staining of uninjected oocytes and oocytes injected 60 h previously with 1.2 ng CK5-GC. The same area of the animal cortex is shown in each case. The antibody CK7 is specific for XCK1(8)⁵ and stains a cortical network of filaments in the same plane as the cortical pigment granules in controls (**c**). XCK1(8) mRNA-depleted oocytes have no visible filaments (**d**). **e**, **f**, Whole-mount staining of blastula (stage 8) embryos with CK7 antibody after injection 20 h earlier of 1 ng AMD7 (**e**) or 1 ng CK5-GC (**f**). Notice the short and disorganized filaments present in **f**.

METHODS. **a**, **b**, Two-dimensional gels were carried out following the protocol



described previously¹⁸. Fixed and dried gels were exposed to pre-flash Fuji X-ray film at -80 °C. The staining method of Klymkowsky²⁰ was used. Before mounting in cavity slides, embryos were incubated for several hours in 0.01% eriochrome black to reduce autofluorescence of yolk. The antibody used was affinity-purified CK7 at 10 µg ml⁻¹.

FIG. 3 The abnormalities caused by XCK1(8) mRNA depletion. **a**, Two mid-blastula stage embryos. The pink embryo is uninjected, whereas the brown one received 1 ng CK5-GC as an oocyte. The individual cells of the brown embryo have a 'cobblestone' appearance. **b**, A cut made in the blastula surface of an XCK1(8)-depleted embryo (brown) did not heal successfully after 40 min compared to the uninjected control (pink). **c**, **d**, **e**, Embryos from experiment 3 in Table 1. **c**, Embryos derived from oocytes injected with 3 ng CK5-GC exogastrulate. Those injected with 3 ng CK5S-GC (**d**) or 3 ng CK5S-GC + 3 ng CK5-GC (**e**) did not show this effect. **METHOD**. Full-grown defolliculated oocytes were injected with oligos in the equatorial region, incubated for 6 h, matured with 1 μ M progesterone and fertilized as described previously⁸. Oocytes were stained using vital dyes to distinguish experimental batches from each other and from the host eggs. Fertilized eggs were sorted, dejellied and allowed to develop. Cuts were made with tungsten needles after removing the vitelline membrane.



(Table 1). Finally, we tested the capacity of the surface cells of the blastulae to repair a small cut in the animal hemisphere. Control embryos healed almost completely within 40 min, whereas cuts in the cytokeatin-depleted embryos failed to heal (Fig. 3b).

These findings are important for three reasons. First, we have demonstrated that the antisense oligo knock-out approach can be used to make 'maternal mutants' in *Xenopus*. To our knowledge, this is the first demonstration of an embryonic phenotype due to such a knock-out. The specificity of the effect is shown by the normal development of embryos injected with irrelevant or sense oligos or with a combination of sense plus antisense CK5-GC oligos. In addition, we have tried to show specificity by overexpression of the normal mRNA after oligo injection. However, we have not yet been able to routinely achieve expression of full-length cytokeatin protein within the time frame of the experiment. Second we have shown that the maternal store of the mRNA for cytokeatin filaments is required in embryogenesis and that the maternal store of protein is itself insufficient for normal development. Third we have provided evidence for the function of keratin filaments in early embryogenesis. The primary effect of the loss of cytokeatin filaments is the inability of the blastula superficial cells to flatten and form an integrated epithelial shell. The failure of cytokeatin-depleted embryos to heal wounds or gastrulate successfully may be explained in two ways. The simplest model is that loss of cytokeatin reduces mechanical rigidity, without which the outer cell layer cannot resist the mechanical forces of the cell movements of gastrulation. This loss of mechanical strength has been suggested as a basis of the skin defects caused by cytokeatin

mutations both in transgenic mice⁹ and in naturally occurring human hereditary conditions^{10,11}. A second possibility is that cytokeatin depletion results in reduced basal and lateral epithelial cell adhesion. Loss of adhesion could result from a quantitative reduction in the amount of surface area available for adhesion in rounded compared to flattened, surface cells. In this regard it will be of some interest to study the interactions of the cytokeatin networks with their associated adhesive junctions in control and cytokeatin-depleted embryos. □

Received 18 February; accepted 21 April 1992.

1. Torpey, N., Heasman, J. & Wylie, C. C. *Development* **110**, 1185-1195 (1990).
2. Franz, J. K. & Franke, W. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6475-6479 (1986).
3. Herman, H., Fouquet, B. & Franke, W. W. *Development* **105**, 279-298 (1989).
4. Klymkowsky, M. W. & Maynell, L. A. *Dev. Biol.* **134**, 479-485 (1989).
5. Torpey, N., Heasman, J. & Wylie, C. C. *J. Cell Sci.* **101**, 151-160 (1992).
6. Newport, J. & Kirschner, M. *Cell* **30**, 687-696 (1982).
7. Dagle, J. M., Walder, J. A. & Weeks, D. L. *Nucleic Acids Res.* **18**, 4751-4757 (1990).
8. Holwill, S. et al. *Development* **100**, 735-743 (1987).
9. Vassar, R., et al. *Cell* **64**, 365-380 (1991).
10. Coulombe, P. A. et al. *Cell* **66**, 1301-1311 (1991).
11. Lane, E. B. et al. *Nature* **356**, 244-246 (1992).
12. Gurdon, J. et al. *Cell* **41**, 913-922 (1985).
13. Sambrook, J. et al. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
14. Lu, X. & Lane, E. B. *Cell* **62**, 681-696 (1991).
15. Domenjoud, L. et al. *Exp. Cell Res.* **179**, 352-361 (1988).
16. Knapp, A. C. & Franke, W. W. *Cell* **59**, 67-79 (1989).
17. Kulesh, D. A. & Oshima, R. G. *Molec. cell. Biol.* **8**, 1540-1550 (1988).
18. Kulesh, D. A. et al. *Molec. cell. Biol.* **9**, 1553-1565 (1989).
19. Tang, P. et al. *Development* **103**, 279-287 (1988).
20. Klymkowsky, M. W., Maynell, L. A. & Polson, A. G. *Development* **100**, 543-557 (1987).

ACKNOWLEDGEMENTS. This work was supported by the Wellcome Trust. We thank K. Goldstone for technical assistance and C. French-Constant and C. Sharpe for their helpful comments on the manuscript.

Regulation of deactivation of photoreceptor G protein by its target enzyme and cGMP

Vadim Yu. Arshavsky* & M. Deric Bownds*†

* Laboratory of Molecular Biology, † Neuroscience Training Program, Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706, USA

THE photoreceptor G protein, transducin, is one of the class of heterotrimeric G proteins that mediates between membrane receptors and intracellular enzymes or ion channels. Light-activated rhodopsin catalyses the exchange of GDP for GTP on multiple transducin molecules. Activated transducin then stimulates cyclic GMP phosphodiesterase by releasing an inhibitory action of the phosphodiesterase γ -subunits. This leads to a decrease in cGMP levels in the rod, and closure of plasma membrane cationic channels gated by cGMP¹⁻⁴. In this and other systems, turn-off of the response requires the GTP bound to G protein to be hydrolysed by an intrinsic GTPase activity⁵⁻⁷. Here we report that the interaction of transducin with cGMP phosphodiesterase, specifically with its γ -subunits, accelerates GTPase activity by several fold. Thus the γ -subunits of the phosphodiesterase serve a function analogous to the GTPase-activating proteins that regulate the class of small GTP-binding proteins. The acceleration can be partially suppressed by cGMP, most probably through the non-catalytic cGMP-binding sites of phosphodiesterase α and β -subunits. This cGMP regulation may function in light-adaptation of the photoresponse as a negative feedback that decreases the lifetime of activated cGMP phosphodiesterase as light causes decreases in cytoplasmic cGMP.

We reconstituted photoreceptor membranes which retained transducin but were depleted of cGMP phosphodiesterase (PDE), with purified PDE or its γ -subunits (PDE γ). Photoreceptor membranes containing transducin tightly bound to bleached rhodopsin were obtained by illuminating the membranes in the absence of nucleotides⁸. More than 80% of the transducin remains bound during multiple hypotonic washes that remove more than 98% of the PDE from the membranes. This transducin will turnover if GTP is added: the α -subunit of transducin binds GTP, dissociates from rhodopsin, and slowly hydrolyses GTP to GDP with an intrinsic GTPase activity. It is this GTPase activity that can be altered if the membranes are first permitted to rebind PDE, by preincubating them with PDE purified from toad rod outer segments to more than 98% homogeneity (Fig. 1). Because the γ -subunit of PDE is a primary site of PDE interaction with transducin⁹, a similar reconstitution with membranes was also done using bovine recombinant PDE γ (ref. 10). To measure the rate of transducin GTPase activity we started the reaction by adding GTP in an amount less than the transducin present in the sample, so that a single turnover of GTP hydrolysis on each activated transducin molecule occurred. The binding of GTP with transducin under the conditions used here occurs at time zero and the subsequent monitoring of inorganic phosphate (P_i) formation reflects the time course of a single synchronized turnover of transducin-bound GTP hydrolysis^{11,12}. The rate constant of the GTPase reaction can be obtained from an exponential fitting of the data.

The rate of the transducin GTPase reaction can be significantly accelerated both by PDE and by PDE γ (Fig. 1). The bottom curve displays the basal GTPase rate of transducin bound to the membranes (0.034 ± 0.003 turnovers per second, $n=6$). Addition of PDE or PDE γ accelerates the rate until saturation (0.143 ± 0.02 turnovers per second, $n=4$ for PDE γ and 0.14 ± 0.019 turnovers per second, $n=3$ for PDE) is observed at $\sim 0.3 \mu\text{M}$ PDE or $\sim 0.6 \mu\text{M}$ PDE γ . Because each PDE molecule contains two γ -subunits¹³, we conclude that the extent of

GTPase acceleration in both experiments is determined by the concentration of PDE γ . Figure 2 demonstrates that this acceleration of GTPase activity is slowed down by cGMP (0.065 ± 0.007 turnovers per second; $n=3$), and that this slowing requires the presence of the α - and β -subunits of PDE which contain catalytic and non-catalytic cGMP-binding sites¹⁴⁻¹⁶. The bottom curve shows that cGMP has no effect on membrane-bound transducin in the absence of PDE. It also does not alter the accelerated GTPase rate caused by addition of PDE γ (top curve). It is only if PDE is present as the $\alpha\beta\gamma_2$ heterotetramer that cGMP suppresses the accelerated GTPase activity. Two observations¹² suggest that the cGMP-sensitivity is conferred through the non-catalytic binding sites located on the PDE α - and β -subunits. The half-saturation of the cGMP effect on transducin GTPase in a suspension of disrupted rod outer segments is very close to the reported value of the dissociation constant (K_D) for the PDE non-catalytic cGMP-binding sites¹⁴ and ~ 100 -fold lower than the reported Michaelis constant (K_M) values associated with the catalytic site of amphibian PDE¹⁷. Further, cAMP can bind to and be hydrolysed by the catalytic site, but cannot substitute for cGMP at the site that regulates GTPase.

Our findings have several implications for understanding photoreception. It has been a paradox that the rate of transducin GTPase measured *in vitro*^{18,19} is too slow to explain the relatively

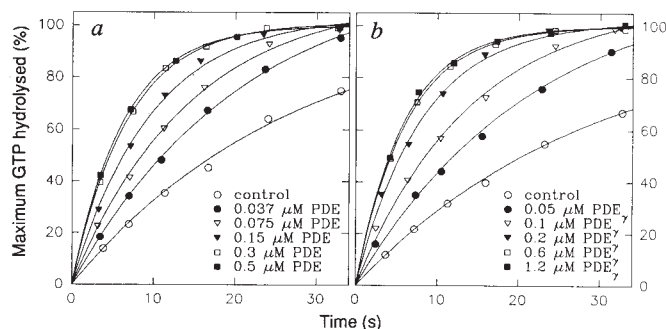


FIG. 1 Acceleration of the transducin GTPase reaction by PDE (a) and recombinant PDE γ (b). Photoreceptor membranes (20 μM rhodopsin) with more than 98% of their PDE extracted but containing $\sim 1.6 \mu\text{M}$ transducin were preincubated with purified PDE or recombinant PDE γ at the indicated concentrations. The reaction was started by adding $0.2 \mu\text{M}$ [γ -³²P]GTP and terminated by perchloric acid. The time course of γ -³²P_i liberation by the GTPase reaction is shown; in all curves 100% hydrolysis corresponds to complete hydrolysis of the $0.2 \mu\text{M}$ GTP added. The data for test membranes alone and for the highest concentrations of PDE or PDE γ are approximated by single exponentials. More complicated kinetics are observed for non-saturating concentrations of PDE and PDE γ and the corresponding lines are hand drawn. The data are taken from one of three (a) or four (b) similar experiments.

METHODS. The isolation of toad rod outer segments and measurements of single-turnover GTPase activity of transducin was as described in ref. 12. All experiments were done in buffer A, which contains 100 mM potassium isethionate 10 mM sodium isethionate, 5 mM MgCl₂, 1 mM dithiothreitol and 15 mM HEPES, pH 7.8. Test membranes retaining >80% of their transducin and depleted by >98% of their PDE were obtained by washing bleached rod outer segments twice with buffer A and twice with buffer B: 5 mM Tris-HCl, pH 8.0, 0.5 mM EDTA and 1 mM dithiothreitol. To prepare purified PDE, non-bleached rod outer segments were washed twice with buffer A and three times with 5 mM Tris-HCl, pH 8.0, containing 5 mM MgCl₂ and 1 mM dithiothreitol. Membranes were then bleached to achieve tight binding of most transducin with rhodopsin, and PDE was extracted by buffer B. After concentrating the extract with a Centricon-30 cartridge (Amicon), PDE was purified by a gel-filtration on a Superose-6 column (Pharmacia) equilibrated with buffer B containing 10 mM NaCl. The only detectable impurity in the PDE fraction was ~ 1 -2% of transducin's α -subunit which was removed by subsequent chromatography on a protein-A-Superose column (Pharmacia) with immobilized anti-transducin antibodies 4A (ref. 30). SDS-polyacrylamide gel electrophoresis of this preparation showed three bands of 92, 87 and 14K corresponding to PDE α -, β - and γ -subunits. Purity was >98%.

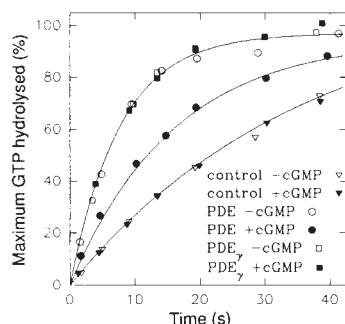


FIG. 2. cGMP reverses transducin GTPase acceleration by PDE but not PDE γ . Single-turnover GTPase measurements were done as described in the legend to Fig. 1. Subsaturing concentrations of PDE (0.2 μ M) and PDE γ (0.4 μ M) which caused the same extent of GTPase acceleration were used. cGMP was added at a saturating concentration of 100 μ M where indicated. To prevent cGMP hydrolysis by PDE during the experiment the PDE inhibitor Zaprinast (100 μ M; May & Baker) was added to all samples. Zaprinast itself did not influence the GTPase reaction (not shown). All the data are taken from one of three similar experiments and approximated by single exponents revealing GTPase rates of 0.031 s^{-1} for test membranes alone, 0.063 s^{-1} for the membranes in the presence of both PDE and cGMP and 0.13 s^{-1} for all other conditions.

fast termination of PDE activation observed under similar conditions²⁰. More recent work indicates that transducin GTPase can be faster under more physiological conditions^{11,21–23}, but the mechanism of GTPase acceleration has remained unclear. The data of Fig. 1 show that PDE itself serves as a GTPase-activating factor. The maximal GTPase rate observed in this reconstitution study ($\sim 0.15 s^{-1}$) is still about 10-fold slower than the rate of the recovery from a photoresponse. But a more rapid rate ($>0.6 s^{-1}$) is observed in suspensions of disrupted rod outer segments¹² for the fast component of GTPase suppressed by micromolar concentrations of cGMP. Our study allows us to conclude that this faster GTPase is a property of that transducin which activates PDE, and thus the extent of PDE-dependent GTPase acceleration is higher in rod outer segment suspensions than in reconstituted membranes. More recent data (V.Y.A. *et al.*, manuscript in preparation) shows that further concentration of rod outer segment suspensions ($>100 \mu$ M rhodopsin) increases GTPase rates by at least twofold, close to the turn-off time of the photoresponse.

The data shown in Fig. 2 indicate a feedback mechanism in retinal rods based on cGMP-dependent regulation of the lifetime of activated PDE. Such a mechanism might function during rod background adaptation, when the duration and light sensitivity of the photoresponse is diminished^{24,25}. A reasonable model is that background light depletes intracellular cGMP levels, causing dissociation of cGMP from the non-catalytic binding sites on PDE. This would accelerate the GTPase activity that terminates each PDE activation event, leading to a faster and/or smaller photoresponse. Such a mechanism might work in parallel with the known calcium feedback regulation of adaptation⁴.

The regulation of GTP-binding protein GTPase activity by an effector described here, although not previously described for a heterotrimeric G protein, has been documented extensively for several other classes of GTP-binding proteins (for example ref. 26). It is observed for elongation and initiation factors whose GTPase activity is enhanced by ribosomes. The class of small GTP-binding proteins including the product of proto-oncogene *ras* interact with GTPase-activating proteins (GAPs) that may also be their effectors²⁷. The intrinsic GTPase of the heterotrimeric signal-transducing G proteins is considerably more rapid than that of the small GTP-binding proteins (for example refs 5–7), but still in several systems such as photoreception^{2,3}, olfaction²⁸ and muscarinic receptor-induced potassium channel regulation²⁹ it has seemed to be too slow to explain the

rapid on-off cycle of the relevant effectors. Because acceleration has now been associated with the effector enzyme in the photo-receptor, it is relevant to search for similar mechanisms in other systems using heterotrimeric G proteins. □

Received 4 February; accepted 8 April 1992.

1. Stryer, L. *A. Rev. Neurosci.* **9**, 87–119 (1986).
2. Liebman, P. A., Parker, K. R. & Dratz, E. A. *A. Rev. Physiol.* **49**, 765–791 (1987).
3. Chabre, M. & Deterre, P. *Eur. J. Biochem.* **179**, 255–266 (1989).
4. Pugh, E. N. & Lamb, T. D. *Vision Res.* **30**, 1923–1948 (1990).
5. Taylor, C. W. *Biochem. J.* **272**, 1–13 (1990).
6. Iyengar, R. & Birnbaumer, L. (eds) *G Proteins* (Academic, San Diego, 1990).
7. Simon, M. I., Strathmann, M. P. & Gautam, N. *Science* **252**, 802–808 (1991).
8. Kuhn, H. *Nature* **283**, 587–589 (1980).
9. Hurley, J. B. & Stryer, L. *J. Biol. Chem.* **257**, 11094–11099 (1982).
10. Brown, R. L. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4922–4926 (1989).
11. Arshavsky, V. Yu., Antoch, M. P., Lukjanov, K. A. & Philippov, P. P. *FEBS Lett.* **250**, 353–356 (1989).
12. Arshavsky, V. Yu., Gray-Keller, M. P. & Bownds, M. D. *J. Biol. Chem.* **266**, 18530–18537 (1991).
13. Deterre, P., Bigay, J., Forquet, F., Robert, M. & Chabre, M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2424–2428 (1988).
14. Yamazaki, A., Sen, I., Bitensky, M., Casnellie, J. E. & Greengard, P. *J. Biol. Chem.* **255**, 11619–11624 (1980).
15. Charbonneau, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 288–292 (1990).
16. Li, T., Volpp, K. & Applebury, M. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 293–297 (1990).
17. Robinson, P. R., Kawamura, S., Abramson, B. & Bownds, M. D. *J. gen. Physiol.* **76**, 631–645 (1980).
18. Fung, B. K.-K., Hurley, J. B. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 152–156 (1981).
19. Baehr, W., Morita, E., Swanson, R. & Applebury, M. L. *J. Biol. Chem.* **257**, 6452–6460 (1982).
20. Sitarumayya, A. & Liebman, P. A. *J. Biol. Chem.* **258**, 12106–12109 (1983).
21. Dratz, E. A., Lewis, J. W., Schaechter, L. E., Parker, K. R. & Kliger, D. S. *Biochem. biophys. Res. Commun.* **146**, 379–386 (1987).
22. Wagner, R., Ryba, N. & Uhl, R. *FEBS Lett.* **234**, 44–48 (1988).
23. Vuong, T. M. & Chabre, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9813–9817 (1991).
24. Fain, G. L. *J. Physiol.* **261**, 71–101 (1976).
25. Baylor, D. A., Lamb, T. D. & Yau, K.-W. *J. Physiol.* **288**, 589–611 (1979).
26. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **349**, 117–127 (1991).
27. Hall, A. *Cell* **61**, 921–923 (1990).
28. Breer, H., Boekhoff, I. & Tareilus, E. *Nature* **345**, 65–68 (1990).
29. Breitwieser, G. E. & Szabo, G. *J. gen. Physiol.* **91**, 469–493 (1988).
30. Witt, P. L., Hamm, H. E. & Bownds, M. D. *J. gen. Physiol.* **84**, 251–263 (1984).

ACKNOWLEDGEMENTS. We thank M. P. Gray-Keller for participating in the preliminary stage of this work, P. D. Calvert and A. Yamazaki for discussion, T. F. J. Martin and A. E. Ruoho for reading the manuscript, C. L. Dumke, M. Vidal and O. Gaide for technical support, L. Stryer and R. L. Brown for recombinant PDE, and H. E. Hamm for antibodies 4A. This work was supported by the NIH.

Aminopeptidase N is a major receptor for the enteropathogenic coronavirus TGEV

Bernard Delmas*, Jacqueline Gelfi*, René L'Haridon*, Lotte Katrine Vogel†, Hans Sjöström†, Ove Norén† & Hubert Laude*‡

* Unité de Virologie et Immunologie Moléculaires, INRA, Domaine de Vilvert, 78350 Jouy-en-josas, France
† Department of Biochemistry C, Panum Institute, DK 2200 Copenhagen N, Denmark

CORONAVIRUSES, like many animal viruses, are characterized by a restricted host range and tissue tropism¹. Transmissible gastroenteritis virus (TGEV), a major pathogen causing a fatal diarrhoea in newborn pig, replicates selectively in the differentiated enterocytes covering the villi of the small intestine². To investigate the molecular determinants of the infection, we characterized the surface molecule used by the virus for binding and entry into host cells. Here we report that aminopeptidase N, an ectoenzyme abundantly expressed at the apical membrane of the enterocytes, serves as a receptor for TGEV. Monoclonal antibodies were selected for their ability to block infection by TGEV of porcine cell lines. They recognized a brush-border membrane protein of *M_r* 150K, which was identified as aminopeptidase N by amino-terminal sequencing. Two lines of evidence supported the view that the peptidase itself

‡ To whom correspondence should be addressed.

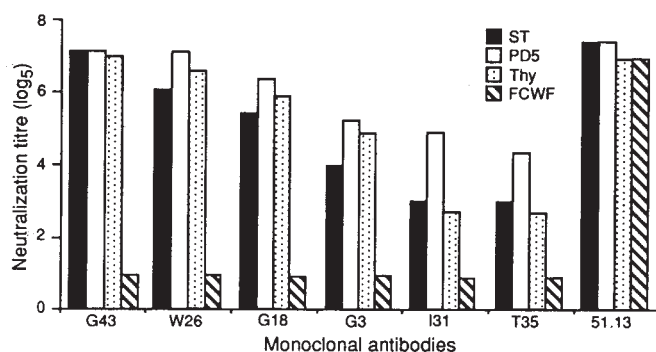


FIG. 1 Characterization of anti-receptor monoclonal antibodies. The neutralization activity of six anti-ST cell antibodies (G43 and T35) was measured in four cell systems permissive to TGEV. The neutralizing titre is expressed as the last of serial fivefold dilutions protecting the monolayer against the viral cytopathic effect (CPE). Monoclonal antibody 51.13 is directed against TGEV spike protein S (ref. 17). ST, PD5 and Thy are porcine testis, kidney and thyroid cell lines, FCWF is a feline cell line.

METHODS. BALB/c mice were immunized by intraperitoneal injection of 5×10^7 intact ST cells three times at 1-month intervals and boosted by injection of 180 μ g ST membranes. The supernatant from 800 hybridoma clones prepared from spleen cells was tested for neutralizing activity using a standard microassay¹⁷ except that the cells were pre-incubated for 2 h with the antibodies. Six positive hybridomas were sub-cloned then amplified. IgGs from ascites fluid were used as a source of antibodies.

acts as a receptor. First, virions bound specifically to aminopeptidase N that was purified to homogeneity. Second, recombinant expression of aminopeptidase N conferred infectivity by TGEV to an otherwise non-permissive cell line.

To obtain monoclonal antibodies against the TGEV receptor, hybridomas were prepared from a mouse immunized with ST cells, a swine testis cell line highly susceptible to TGEV. Several of the resulting antibodies exhibited, in three different porcine cell systems, a blocking activity comparable to that of a high-titre neutralizing anti-TGEV antibody (Fig. 1). By contrast, no significant protection by the antibodies was observed after virus challenge in a feline cell system, or towards irrelevant viruses (group A bovine rotavirus or vesicular stomatitis virus; data not shown). Therefore, the selected antibodies had the characteris-

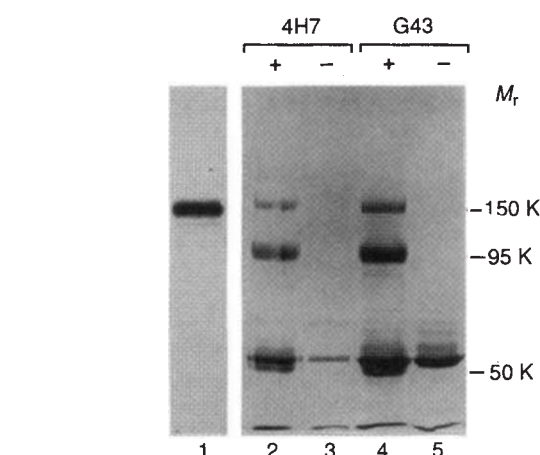


FIG. 2 Polypeptides recognized by anti-receptor monoclonal antibodies. Lane 1, 35 S-labelled polypeptides from ST-cell extracts immunoprecipitated by antibody G43. Lanes 2-5, polypeptides from pig intestinal brush-border membrane immunoprecipitated by the anti-aminopeptidase N (APN) antibody, 4H7 (ref. 4) or by antibody G43. Antibodies were incubated in the presence (+) or absence (-) of solubilized brush-border membrane.

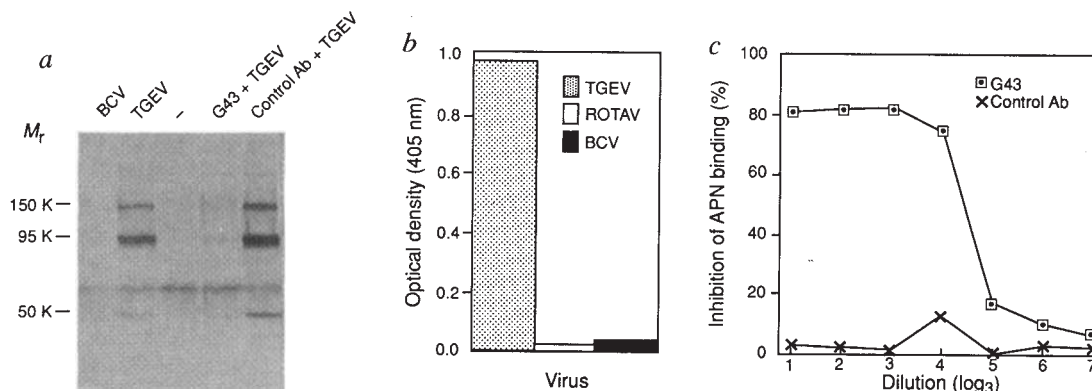
METHODS. ST cells were labelled by incubation for 6 h with 100 μ Ci ml^{-1} 35 S-methionine (Amersham) after a 2-h methionine depletion. Intestinal microvillar membranes were prepared as described¹⁸. Cells or microvillar membranes were solubilized in immunoprecipitation buffer¹⁷. Immunoprecipitated polypeptides were resolved in 8% SDS-PAGE then fluorographed (lane 1) or stained with Coomassie blue (lanes 2-5).

tics expected for antibodies recognizing a major TGEV receptor.

The monoclonal antibodies all recognized a polypeptide of relative molecular mass 150,000 (150K) in ST cell extracts, together with a faint band interpreted as the mannose-rich intracellular precursor (Fig. 2, lane 1). When using solubilized brush-border membranes from pig small intestine, three major species of 150K, 95K and 50K were coimmunoprecipitated (Fig. 2, lane 4). The first 30 amino acids of the 95K species were determined through N-terminal sequencing: NH_2 -Ala-Lys-Gly-Phe-Tyr-Ile-Ser-Lys-Ala-Leu-Gly-Ile-Leu-Gly-Ile-Leu-Leu-Gly-Val-Ala-Ala-Val-Ala-Thr-Ile-Ile-Ala-Leu-Ser-Val-COOH. This sequence was identical to the N-terminal sequence (minus the first Met) of porcine aminopeptidase N, deduced from the exon I nucleotide sequence³.

FIG. 3 Demonstration of a virus-receptor binding with purified components. *a*, Aminopeptidase N (APN) preincubated with or without purified antibody G43 or a control antibody was incubated in the absence (-) or presence of TGEV or bovine coronavirus (BCV) virions, then centrifuged through a glycerol cushion. The presence of APN-specific bands in the pellets was revealed by western blotting. *b*, Left panel, soluble APN was incubated on plastic dishes coated with TGEV, BCV or rotavirus virions. Bound APN was revealed by immunoassay. Right panel, APN was preincubated with serial dilutions of either G43 or a control antibody and tested for TGEV binding as above.

METHODS. Anchor-free APN was purified by immunoabsorbent chromatography from porcine intestine microvillar membranes after Triton X-100 solubilization and trypsin treatment⁵. The purity of APN polypeptides was confirmed by SDS-PAGE and Coomassie blue staining. TGEV and BCV virions were purified as described¹⁷. Rotavirus virions were purified on a CsCl gradient. *a*, APN (0.2 μ g) was mixed with a 50 μ g virion suspension in cell



culture medium. After 1 h at 37 $^{\circ}$ C, the virions were pelleted through a 10% glycerol cushion by centrifugation at 150,000g for 30 min. One TGEV sample was mixed with APN preincubated with G43 IgGs (200 μ g ml^{-1}). APN bound to virus was revealed by western blotting using rabbit IgGs directed against denatured APN¹⁹ and a peroxidase conjugate. *b*, APN (0.6 μ g) was added to virus-coated wells (1 μ g per well) for 1 h at 37 $^{\circ}$ C after incubation (or mock incubation) with dilutions of G43 IgGs at 200 μ g ml^{-1} . After washes with PBS plus 0.05% Tween 20, bound APN was detected by rabbit IgGs against native APN and a phosphatase conjugate.

FIG. 4 Susceptibility to TGEV of MDCK cells stably expressing porcine APN. **a**, Expression of APN in transfected (lane 1) and non-transfected (lane 2) MDCK clones. G43 antibody immunoprecipitates (lanes 1 and 2) and APN from pig brush-border membrane (lane 3) were analysed by western blotting; **b**, Colorimetric quantification of APN-MDCK and MDCK cells after an infection or mock-infection (–) by TGEV; in two assays, G43 antibody or bestatin were added before infection. The data are given as mean values + s.d.m. ($n=8$). **c**, Synthesis of TGEV-specific polypeptides S, M and N (ref. 17) in infected APN-MDCK (lane 1) and MDCK cells (lane 2).

METHODS. cDNA was synthesized from poly(A)⁺ RNA isolated from pieces of porcine intestine. cDNAs larger than two kilobases were cloned into the *EcoRI* site of the λ ZapII bacteriophage (Stratagene). A full-length cDNA clone was isolated from the cDNA library by probing with the *PstI* fragment (–55, +543) derived from exon 1 of the porcine *APN* gene³. The clone was sequenced using the dideoxy method. A *BamHI*–*BglII* fragment (3,280 base pairs) covering the complete *APN* open reading frame was subcloned downstream of the ubiquitin promoter into the *BamHI* site of pTEJ4 expression vector²⁰. MDCK cells were cotransfected with this construct and pSV2neo. Cell clones resistant to the neomycin analogue G418 were selected and assayed for APN expression. Material immunoprecipitated with G43 antibody from the Triton X-100 solubilized cell lysates was analysed by western blotting as described in

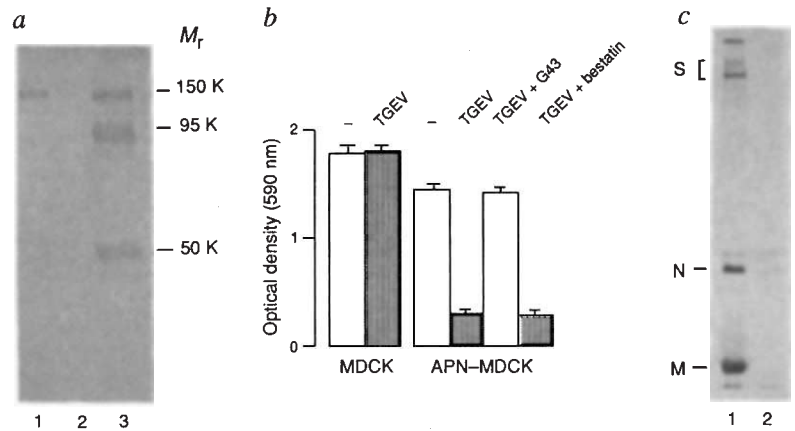


Fig. 3 legend. Colorimetric assays were done 16 h after infection at a multiplicity of five plaque-forming units¹⁷. Monolayers were fixed and stained with a crystal violet solution. The dye incorporated in cells surviving the viral CPE was measured by optical absorbance after solubilization in acetic acid²¹. Cells were incubated in the presence of G43 antibody or bestatin (1 mM) from 1 h before infection. Immunoprecipitation of ³⁵S-methionine-labelled intracellular polypeptides was as described for Fig. 2.

Further evidence that the anti-TGEV-receptor antibodies recognized aminopeptidase N was obtained by showing that (1) an antibody raised against rabbit aminopeptidase N⁴ reacted with the same three polypeptides in brush-border membrane preparations (Fig. 2, lane 2): 95K and 50K, corresponding to the B (amino) and C (carboxy) subunits of the pig aminopeptidase, and 150K, uncleaved aminopeptidase⁵; (2) the immunoprecipitated material hydrolysed leucine *p*-nitroanilide, a chromogenic substrate specific for aminopeptidase (ref. 6; data not shown).

Two experiments were designed to demonstrate any direct association between aminopeptidase N and the virus. First, soluble aminopeptidase N was centrifuged after incubation in the presence of virions (Fig. 3a). Aminopeptidase N-specific bands were recovered with pelleted TGEV virions only. Second, when the aminopeptidase was incubated in the presence of adsorbed virions (Fig. 3b), it bound to TGEV and not the other enteric viruses. In both assays, earlier incubation with an antibody against aminopeptidase N reduced the binding considerably. Because the two components were purified to homogeneity, it was concluded that the interaction between the aminopeptidase and TGEV occurs in the absence of any other cellular protein.

The gene encoding aminopeptidase N (*APN*) was expressed in non-permissive cells to see whether this would confer them with the capacity to bind TGEV. A pig intestine complementary DNA library was screened by use of a homologous DNA probe derived from the 5' end of *APN* gene. A full-length cDNA copy was cloned and contained an open reading frame of 2,889 nucleotides encoding a polypeptide 79% identical to human aminopeptidase (data not shown). MDCK cell clones stably transformed with the porcine *APN* cDNA expressed a polypeptide of 150K which reacted with antibodies against aminopeptidase N (Fig. 4a). The aminopeptidase activity⁶ of the transfected clones was about 40-fold higher compared with non-transfected clones. On viral challenge, all of the three independent clones tested seemed to be fairly susceptible to TGEV infection, as proved by extensive destruction of the infected monolayers and synthesis of the viral structural polypeptides (Fig. 4b, c). Earlier incubation with an antibody specific for aminopeptidase N prevented the appearance of viral cytopathic effect. These results show that aminopeptidase N was the only porcine protein necessary to confer susceptibility on canine kidney cells naturally resistant to TGEV. Moreover, the protease function

of the molecule did not seem to be involved because it was blocked by bestatin, an inhibitor of aminopeptidase, without preventing the infection (Fig. 4b).

So far, defined receptors include molecules that belong to the immunoglobulin superfamily, such as CD4 for HIV⁷, ICAM-1 for rhinovirus⁸, poliovirus receptor⁹ and a carcinoembryonic antigen for murine hepatitis coronavirus¹⁵, and also an amino-acid transporter for murine leukaemia retroviruses¹¹. Our study provides strong evidence that porcine aminopeptidase N serves as a receptor for an enveloped RNA virus, TGEV. This emphasizes the diversity of the membrane-bound proteins that viruses subvert for gaining entry into cells.

Aminopeptidase N is a well documented ectoenzyme that binds to the membrane through an N-terminal segment^{5,12,13}. Human aminopeptidase N is identical to CD13, a surface antigen of many myeloid cells¹⁴. It is a zinc-binding protease that catalyses the removal of N-terminal, preferentially neutral residues from peptides. It is expressed in many tissues at different levels¹⁵, the highest activity being found in the small intestinal mucosa, where the aminopeptidase represents about 8% of the protein content of the apical membrane of the differentiated enterocytes, and in the kidney brush border. It is also expressed to a lesser extent in liver, lung and colon, where the virus does replicate, but without causing the specific histopathological damage seen in the small intestine¹⁶. In the intestine, the distribution of the receptor and the site of multiplication of TGEV are thus strikingly correlated. This argues for a pivotal role of aminopeptidase N/CD13 in determining the tissue tropism of TGEV. Investigating the nature of the virus interaction with aminopeptidase N could provide a rationale for the design of an antiviral strategy against TGEV and related infections. □

Received 5 February; accepted 2 April 1992.

- Siddell, S., Wege, H. & Ter Meulen, V. *J. gen. Virol.* **64**, 761–776 (1983).
- Pensaert, M., Haelterman, E. O. & Burstein, T. *Arch. Gesamte. Virusforsch.* **31**, 321–334 (1970).
- Olsen, J., Sjöström, H. & Norén, O. *FEBS Lett.* **251**, 275–281 (1989).
- Gorvel, J. P. et al. *J. Cell Biol.* **108**, 2193–2200 (1989).
- Sjöström, H. et al. *Eur. J. Biochem.* **88**, 503–511 (1978).
- Kramers, M. T. C. & Robinson, G. B. *Eur. J. Biochem.* **99**, 345–351 (1979).
- Maddon, P. J. et al. *Cell* **47**, 333–348 (1986).
- Greve, J. M. et al. *Cell* **56**, 839–847 (1989).
- Mendelsohn, C. L., Wimmer, E. & Racaniello, V. R. *Cell* **56**, 855–865 (1989).
- Dveksler, G. S. et al. *J. Virol.* **65**, 6881–6891 (1991).
- Kim, J. W., Closs, E. I., Albritton, L. M. & Cunningham, J. M. *Nature* **352**, 725–728 (1991).
- Maroux, S., Louvard, D. & Baratti, J. *Biochim. biophys. Acta* **321**, 282–295 (1973).
- Olsen, J. et al. *FEBS Lett.* **238**, 307–314 (1988).

14. Look, A. T., Ashmun, R. A., Shapiro, L. H. & Peiper, S. C. *J. clin. Invest.* **83**, 1299–1307 (1989).
15. Kenny, A. J. & Maroux, S. *Physiol. Rev.* **62**, 91–128 (1982).
16. Haeltterman, E. O., *J. Am. vet. med. Assoc.* **160**, 534–540 (1972).
17. Laude, H. et al. *J. gen. Virol.* **67**, 119–130 (1986).
18. Hauser, H., Howell, K., Dawson, R. M. C. & Bowyer, D. E. *Biochim. biophys. Acta* **602**, 567–577 (1980).
19. Sjöström, H. & Norén, O. *Eur. J. Biochem.* **122**, 245–250 (1982).
20. Johansen, T. H., Schiller, M. S., Tolstoy, S. & Schwartz, T. W. *FEBS Lett.* **267**, 289–294 (1990).
21. Kueng, W., Silber, E. & Eppenberger, U. *Analyt. Biochem.* **182**, 16–19 (1989).

ACKNOWLEDGEMENTS. We thank R. Christon for microvillar preparations, J. C. Huet for N-terminal sequencing, S. Maroux for anti-APN antibodies, and D. Rasschaert and M. Breumont for discussion. B.D. acknowledges the support of the Commission of the European Communities. L.K.V., H.S. and O.N. are members of the Biomembrane Research Center, Aarhus University and are supported by a grant from the Lundbeck foundation.

Human aminopeptidase N is a receptor for human coronavirus 229E

Curtis L. Yeager*, Richard A. Ashmun†‡, Richard K. Williams*, Christine B. Cardellicchio*, Linda H. Shapiro†, A. Thomas Look†‡§ & Kathryn V. Holmes*||

* Department of Pathology, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, Maryland 20814–4799, USA
 † Departments of ‡ Hematology/Oncology and ‡ Tumor Cell Biology, St Jude Children's Research Hospital, 332 North Lauderdale, Memphis, Tennessee 38105, USA

§ Department of Pediatrics, University of Tennessee College of Medicine, Memphis, Tennessee 38163, USA

HUMAN coronaviruses (HCV) in two serogroups represented by HCV-229E and HCV-OC43 are an important cause of upper respiratory tract infections¹. Here we report that human aminopeptidase N, a cell-surface metalloprotease on intestinal, lung and kidney epithelial cells^{2–5}, is a receptor for human coronavirus strain HCV-229E, but not for HCV-OC43. A monoclonal antibody, RBS, blocked HCV-229E virus infection of human lung fibroblasts, immunoprecipitated aminopeptidase N and inhibited its enzymatic activity. HCV-229E-resistant murine fibroblasts became susceptible after transfection with complementary DNA encoding human aminopeptidase N. By contrast, infection of human cells with HCV-OC43 was not inhibited by antibody RBS and expression of aminopeptidase N did not enhance HCV-OC43 replication in mouse cells. A mutant aminopeptidase lacking the catalytic site of the enzyme did not bind HCV-229E or RBS and did not render murine cells susceptible to HCV-229E infection, suggesting that the virus-binding site may lie at or near the active site of the human aminopeptidase molecule.

To develop a monoclonal antibody against the HCV-229E receptor, we produced hybridomas against deoxycholate-solubilized membrane proteins of two HCV-229E-susceptible human cell lines (WI38 lung fibroblasts and HL60 myeloid leukaemia cells). A monoclonal antibody designated RBS protected WI38 and RD human cell lines from HCV-229E-induced cytopathic effects and protected WI38 cells from virus infection (Fig. 1a–c). RBS pretreatment reduced the number of HCV-229E-infected WI-38 cells at 10 h post-infection by 96%, compared with cells pretreated with control mouse ascites. By contrast, RBS did not inhibit replication of HCV-OC43 in WI38 or RD cells, indicating that the receptor specificities of HCV-OC43 and HCV-229E are different.

Susceptibility to HCV-229E infection in mouse-human somatic cell hybrids depends on a gene located on human chromosome 15 (ref. 6). A promising candidate for the HCV-

TABLE 1 Biological activities of anti-aminopeptidase N monoclonal antibodies and aminopeptidase N inhibitors

	Inhibition of enzyme activity (%)*	Binding to hAPN _{mut} -3T3†	Inhibition of HCV-229E infection‡
Monoclonal antibodies			
WM15	91	—	+
RBS	90	—	+
MY7	42	+	+
Chemical inhibitors§			
Actinonin	100	NA	—
Bestatin	100	NA	—
1,10-Phenanthroline	100	NA	+
2,2'-Dipyridyl	100	NA	+

* The inhibition of hAPN activity was determined as described in the legend to Fig. 2d.

† Binding of antibodies to hAPN_{mut}-3T3 cells was measured by flow cytometry, as outlined in the legend to Fig. 3. The mutant lacks peptidase activity; thus, assays for chemical inhibition were not applicable (NA).

‡ Confluent monolayers of WI38 cells in 96-well plates were pretreated with dilutions of antibodies or inhibitors in medium for 1 h, and then challenged with 1×10^3 p.f.u. per well of HCV-229E. After 1 h of adsorption, the inoculum was removed, and the cells were incubated with fresh medium containing antibodies or inhibitors for 48 h, at which time the monolayers were examined for virus-induced cytopathic effects. Such effects were evident in HCV-229E-infected controls pretreated with normal serum, but not in mock-infected controls. Plus signs, HCV-229E-induced cytopathic effects were inhibited by antibodies up to a dilution of 1:200. All incubations were at 37 °C.

§ Inhibitors were tested at the following concentrations: bestatin, 1 mg ml⁻¹; 1,10-phenanthroline, 1.5 mM; 2,2'-dipyridyl, 2.5 mM; actinonin, 2.7 mM. Antibodies were tested at concentrations that saturated available binding sites in flow cytometric assays.

229E receptor is human aminopeptidase N (hAPN; EC 3.4.11.2), a cell-surface glycoprotein encoded by a gene on bands q25–q26 of human chromosome 15 (ref. 7) and expressed on human lung, renal and intestinal epithelial cells, fibroblasts and nerve synapses^{2–5}. This exopeptidase removes amino-terminal residues to complete the digestion of short peptides in the gut and helps break down neurotransmitter peptides in the brain^{2,3,5,8}. hAPN is identical to CD13, a glycoprotein identified on granulocytes, monocytes and their bone marrow progenitors^{9,10}. Porcine aminopeptidase N is a receptor for transmissible gastroenteritis virus, a porcine coronavirus in the same serogroup as HCV-229E (ref. 11). Because aminopeptidase from humans, pigs and other mammals are structurally similar^{9,12–14}, we investigated whether HCV-229E and RBS would bind specifically to hAPN and whether expression of hAPN by murine cells would make them susceptible to infection with HCV-229E.

Murine NIH3T3 cells transfected with hAPN cDNA in a retroviral vector⁹ (hAPN-3T3) and untransfected NIH3T3 cells were challenged with HCV-229E and HCV-OC43 to determine their susceptibility to virus infection. Although the control NIH3T3 cells were resistant to HCV-229E infection (Fig. 1d), the hAPN-transfected mouse cells were susceptible to infection with this virus (Fig. 1e). By contrast, hAPN-3T3 cells were no more susceptible than NIH3T3 cells to infection with HCV-OC43 (data not shown). Thus, expression of hAPN confers HCV-229E susceptibility, but not HCV-OC43 susceptibility, on murine cells.

We analysed binding of RBS to membrane preparations from hAPN-3T3 or parental NIH3T3 fibroblasts. The antibody bound to membranes of hAPN-3T3 but not to those of NIH3T3 cells (Fig. 2a), indicating that RBS recognized hAPN. Similarly, HCV-229E virions bound more strongly to hAPN-3T3 membranes than to NIH3T3 membranes (Fig. 2b), and RBS competi-

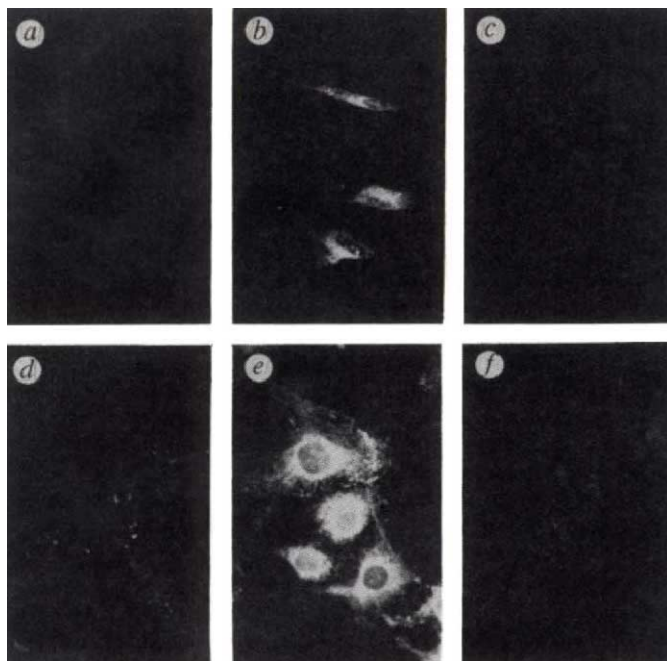


FIG. 1 Inhibition of HCV-229E replication in WI38 cells by anti-receptor monoclonal antibody RBS and susceptibility to HCV-229E of mouse cells expressing normal and mutant hAPN. *a*, Uninfected WI38 cells; *b*, WI38 cells infected with HCV-229E; or *c*, WI38 cells treated with RBS and then challenged with HCV-229E. Three types of murine NIH3T3 cells were challenged with HCV-229E: *d*, native cells; *e*, transformants expressing high levels of hAPN (hAPN-3T3)⁷; and *f*, cells engineered to express a mutant hAPN (hAPN_{mut}3T3)¹¹, which lacks 39 amino acids, including the active site of the enzyme.

METHODS. Hybridoma supernatants (1,624) were tested for the ability to inhibit HCV-229E infection of WI38 cells and one, RBS, was positive. The RBS hybridoma was subcloned three times, and ascites fluid produced by these cells in BALB/c mice was used as the source of antibody. Cells grown on coverslips were pretreated for 1 h at 37 °C with a 1:10 dilution of RBS, control ascites or control medium, and then challenged with 1×10^4 p.f.u. HCV-229E. Cells were incubated at 37 °C for 10 h (*a-c*) or for 18 h (*d-f*) and fixed in cold acetone. Antisera remained in the plates throughout the experiment. Intracellular HCV-229E antigens were detected with goat anti-HCV-229E and rhodamine-labelled rabbit anti-goat immunoglobulin.

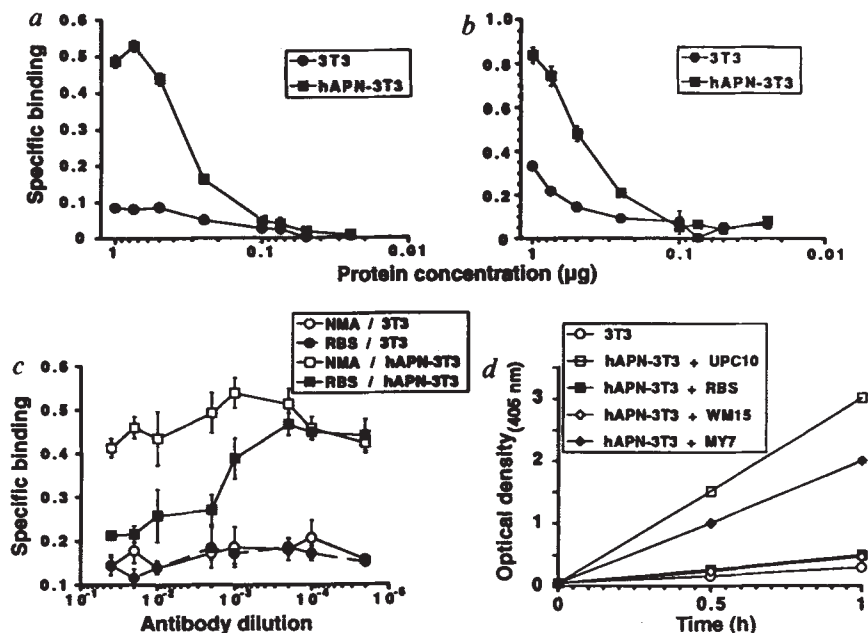
tively inhibited virus binding to hAPN-3T3 membranes (Fig. 2c), suggesting that RBS and HCV-229E may recognize adjacent or overlapping epitopes of the peptidase. Some, but not all, antibodies that bind to extracellular epitopes of hAPN inhibit enzymatic activity¹⁵. We compared the ability of RBS to inhibit the enzymatic activity of hAPN with that of hAPN/CD13-

specific antibodies MY7 and WM15. RBS blocked hAPN activity as efficiently as WM15 and better than MY7 (Fig. 2d).

Flow cytometric results indicated that both RBS and MY7 bound to hAPN-3T3 cells (Fig. 3a, b), but not to untransfected NIH3T3 cells. RBS also immunoprecipitated the mature 150K hAPN glycoprotein and its 130K intracellular precursor¹⁶ from

FIG. 2 Competitive binding of the RBS antibody and HCV-229E to hAPN. Native mouse NIH3T3 cells (3T3) or NIH3T3 transformants expressing high levels of hAPN (hAPN-3T3) were tested. Membrane preparations were used in ELISA assays to measure: *a*, RBS binding; *b*, HCV-229E binding; or *c*, competitive binding of RBS and HCV-229E. *d*, Aminopeptidase N enzymatic activity on the surface of intact cells in the presence of RBS, the hAPN-specific antibodies WM15 or MY7, or the mouse myeloma protein UPC10 (control).

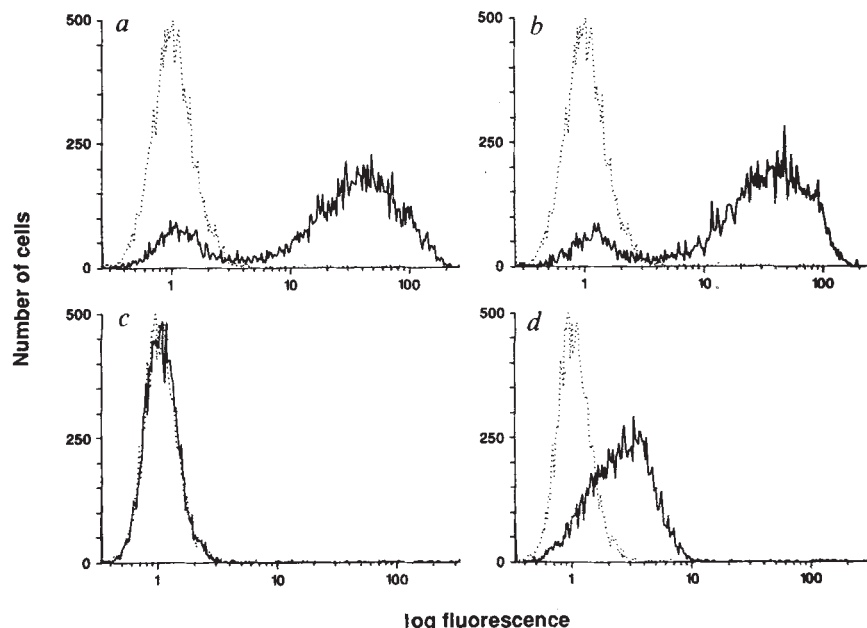
METHODS. Cells were Dounce-homogenized and nuclei were removed by low-speed centrifugation. Cell lysates were pelleted by ultracentrifugation and the membranes resuspended in 10 mM Tris-EDTA with 1% aprotinin. Protein concentrations were determined by Bradford assay (Bio-Rad). Membrane preparations were serially diluted in carbonate buffer, pH 9.6, incubated overnight on Dynatech Immulon 1 microtitre plates, blocked with TBS-Tween buffer (0.15 M NaCl, 0.05 M Tris-HCl, 0.001 M EDTA, 0.05% Tween-20) containing 5% BSA. For detection of antibody binding, 100 μ l of a 1:50 dilution of either RBS or normal mouse ascites in TBS-Tween buffer with 0.1% BSA was added to each well, and peroxidase-labelled goat anti-mouse immunoglobulin (Boehringer Mannheim) diluted 1:1,000 was used to detect bound antibody. For detection of HCV-229E binding, membrane-coated plates were incubated with dilutions of HCV-229E or control medium, followed by goat anti-HCV-229E, biotinylated rabbit anti-goat immunoglobulin and streptavidin-peroxidase (Kirkegaard and Perry). For competitive inhibition of HCV-229E binding by RBS, dilutions of RBS ascites in medium were incubated with membranes on plates before and after the addition of HCV-229E. Incubations were at room temperature for 1 h. Assays were completed by addition of TMB substrate (Kirkegaard and Perry) and read on a Dynatech laboratories MR700 ELISA Reader. Background values obtained with membrane preparations alone were subtracted to give values for specific binding.



For measurements of enzymatic activity on intact cells, 2×10^5 cells were washed in PBS, resuspended in isotonic defined medium (50 mM phosphate buffer, pH 7.4) and incubated at 37 °C for 1 h in an excess of RBS, WM15, MY7 or UPC10 (control mouse myeloma protein). After incubation, and while still at 37 °C, the APN substrate, alanine-*p*-nitroaniline, was added to a final concentration of 6 mM. Samples were periodically removed and chilled to 4 °C to arrest enzymatic activity. After centrifugation, free *p*-nitroaniline present in the cell-free supernatants was detected by absorbance measurement at 405 nm. Degradation of substrate in the absence of cells over the same time period was minimal. All measurements were made in triplicate.

FIG. 3 Flow cytometric analysis of antibody binding to NIH3T3 transfectants. *a*, The hAPN-3T3 cells incubated with either the RBS antibody (solid line) or the control mouse myeloma protein UPC10 (dotted line). *b*, The hAPN-3T3 cells incubated with either the MY7 antibody (solid line) or UPC10 (dotted line). *c*, The hAPN_{mut}-3T3 cells incubated with either the RBS antibody (solid line) or UPC10 (dotted line). *d*, The hAPN_{mut}-3T3 cells incubated with either the MY7 antibody (solid line) or UPC10 (dotted line).

METHODS. Cells (1×10^6) were incubated at 4 °C for 30 min in a titred excess of RBS, MY7 (Coulter), or control mouse myeloma protein (UPC10; Organon-Teknika-Cappel). After being washed twice with cold staining medium (Dulbecco's modified Eagle's medium, supplemented with 10% fetal bovine serum, L-glutamine, 10 mM HEPES, antibiotics and 2 mM sodium azide), the cells were incubated at 4 °C for 30 min in a titred excess of fluoresceinated, affinity-purified goat antiserum to mouse immunoglobulins (Coulter). Washed cells were resuspended in cold staining medium containing 0.25 mM propidium iodide and those that excluded the dye were analysed by flow cytometry.



hAPN-3T3 cells (data not shown). RBS did not bind to NIH3T3 cells transfected with a cDNA encoding a mutant hAPN glycoprotein (hAPN_{mut}-3T3; Fig. 3c) lacking a 39-amino-acid region (amino acids 360–398) which includes the His-Glu-X-X-His sequence that mediates zinc binding and catalytic activity in a series of related metalloproteases¹⁸. MY7 bound to hAPN_{mut}-3T3 cells (Fig. 3d), confirming expression of the mutant protein on the cell surface. But hAPN_{mut}-3T3 cells were not susceptible to HCV-229E infection (Fig. 1f), indicating that both RBS and HCV-229E recognize epitopes that were eliminated or altered in the mutant polypeptide.

We investigated whether the ability of antibodies to block HCV-229E infectivity correlated with their ability to block enzymatic activity or to bind to the mutant hAPN polypeptide. RBS and WM15 both blocked enzyme activity and virus infection, but failed to bind to hAPN_{mut}-3T3 cells (Table 1). By contrast, although MY7 also inhibited HCV-229E infection, it only partially inhibited enzyme activity and recognized the mutant hAPN glycoprotein. Infection of WI38 cells by HCV-229E was not prevented by either actinonin or bestatin, small inhibitory molecules that bind competitively to the active site of hAPN^{19–21}. By contrast, HCV-229E infection was inhibited by 1,10-phenanthroline and 2,2'-dipyridyl, which block aminopeptidase N activity by chelating the zinc required by enzymatically active hAPN molecules¹⁸. These chelators may alter the conformation

of hAPN epitopes, preventing virus attachment. Other compounds that prevent binding of HCV-229E to hAPN may prove useful as drugs against infection with HCV-229E and related coronaviruses.

We have shown that human coronavirus HCV-229E uses human aminopeptidase N as a receptor and that antibodies directed against the active site of this enzyme prevent virus infection of human cells. A serologically unrelated human respiratory coronavirus, HCV-OC43, does not use hAPN as a receptor, and its receptor is unknown. HCV-OC43 is related to murine coronavirus MHV-A59, whose receptor^{22–24} is a member of the carcinoembryonic antigen family of glycoproteins in the immunoglobulin superfamily. Thus, two related coronaviruses, HCV-229E and TGEV, use aminopeptidase N glycoproteins of their normal host species as receptors, whereas two coronaviruses in a different serogroup do not use this aminopeptidase as their receptors. Possibly, coronaviruses related to HCV-229E, such as canine or feline coronaviruses, also use as receptors aminopeptidase or related metalloproteases, such as aminopeptidase A (refs 25, 26) or neutral endopeptidase^{26–30}. Because of their abundance at the apical surfaces of epithelial cells in the alimentary and respiratory tracts, these metalloproteases are promising candidates for receptors of other groups of mammalian viruses or for bacteria or fungi. □

Received 3 February; accepted 2 April 1992.

- McIntosh, K. in *Virology* 2nd edn (eds Fields, D. N. et al.) 857–864 (Raven, New York, 1990).
- Noren, O., Sjöström, H., Danielson, E. M., Cowell, G. M. & Skovbjerg, H. in *Molecular and Cellular Basis of Digestion* (ed. Desnuelle, P.) 335–363 (Elsevier, Amsterdam, 1986).
- Semenza, G. A. *Rev. Cell Biol.* **2**, 255–313 (1986).
- Kenny, A. J. & Maroux, S. *Physiol. Rev.* **62**, 91–128 (1982).
- Matsas, R., Stephenson, S. L., Hryszko, J., Kenny, A. J. & Turnover, A. J. *Biochem. J.* **231**, 445–449 (1985).
- Sakaguchi, A. Y. & Shows, T. B. *Somatic Cell Genet.* **8**, 83–94 (1982).
- Look, A. T. et al. *J. clin. Invest.* **78**, 914–921 (1986).
- Turner, A. J., Hooper, N. M. & Kenny, A. J. in *Mammalian Ectoenzymes* (eds Kenny, A. J. & Turner, A. J.) 211–248 (Elsevier, Amsterdam, 1987).
- Look, A. T., Ashmun, R. A., Shapiro, L. H. & Peiper, S. C. *J. clin. Invest.* **83**, 1299–1307 (1989).
- Look, A. T. et al. in *Leukocyte Typing IV* (eds Knapp, W. et al.) 784–787 (Oxford Univ. Press, Oxford, 1989).
- Delmas, B. et al. *Nature* **357**, 417–420 (1992).
- Olsen, J., Sjöström, H. & Noren, O. *FEBS Lett.* **251**, 275–281 (1989).
- Noren, O. et al. *FEBS Lett.* **259**, 107–112 (1989).
- Watt, V. M. & Yip, C. C. *J. biol. Chem.* **264**, 5480–5487 (1989).
- Ashmun, R. A. & Look, A. T. *Blood* **75**, 462–469 (1990).
- Look, A. T. et al. *J. clin. Invest.* **75**, 569–579 (1985).
- Ashmun, R. A., Shapiro, L. H. & Look, A. T. *Blood* (in the press).

- Vallée, B. L. & Auld, D. S. *Biochemistry* **29**, 5647–5659 (1990).
- Umezawa, H. et al. *J. Antibiotics* **38**, 1629–1630 (1985).
- Rich, D. H., Moon, B. J. & Harbeson, S. J. *med. Chem.* **27**, 417–422 (1984).
- Wilkes, S. H. & Prescott, J. M. *J. biol. Chem.* **260**, 13154–13162 (1985).
- Williams, R. K., Jiang, G. S., Snyder, S. W., Frana, M. F. & Holmes, K. V. *J. Virol.* **64**, 3817–3823 (1990).
- Williams, R. K., Jiang, G. S. & Holmes, K. V. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5533–5536 (1990).
- Dveksler, G. S. et al. *J. Virol.* **65**, 6881–6891 (1991).
- Wu, Q., Lahti, J. M., Air, G. M., Burrows, P. D. & Cooper, M. D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 993–997 (1990).
- Wu, Q., Li, L., Cooper, M. D., Pierres, M. & Gorvel, J. P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 676–680 (1991).
- Devault, A. et al. *EMBO J.* **6**, 1317–1322 (1987).
- Malfroy, B., Kuang, W.-J., Seeburg, P. H., Mason, A. J. & Schofield, P. R. *FEBS Lett.* **229**, 206–210 (1988).
- Letarte, M. et al. *J. exp. Med.* **168**, 1247–1253 (1988).
- Shipp, M. A. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 297–301 (1989).

ACKNOWLEDGEMENTS. We thank H. Nguyen, E. Wingfield, E. Mann and K. Coleman for technical assistance, D. Snyder, C. Dieffenbach, G. Dveksler and D. Wessner for discussions, and J. Gilbert for review.

Crystalline ribonuclease A loses function below the dynamical transition at 220 K

Bjarne F. Rasmussen, Ann M. Stock, Dagmar Ringe, & Gregory A. Petsko

Structural Biology Laboratory, Rosenstiel Basic Medical Sciences Research Center, and Departments of Biochemistry and Chemistry, Brandeis University, Waltham, Massachusetts 02254-9110, USA

WHEN the dynamic properties of many different proteins are plotted as a function of temperature, biphasic behaviour is observed, with a broad transition centred around 220 K. Atomic mean-square displacements from X-ray crystallography¹ and Mössbauer scattering²⁻³ show this behaviour, as do electron transfer rates⁴ and dynamic information from inelastic neutron scattering⁵. Molecular dynamics simulations over a range of temperatures also exhibit a transition at about 220 K: high-temperature atomic fluctuations are dominated by anharmonic collective motions of bonded and nonbonded groups of atoms, but below 220 K the predominant dynamic behaviour is harmonic vibration of individual atoms⁶. Here we show by high-resolution X-ray diffraction that crystalline ribonuclease A does not bind substrate or inhibitor at 212 K but will bind either rapidly at 228 K. Once bound at the higher temperature, inhibitor cannot be washed off after the enzyme is cooled to below the transition temperature. These results suggest that enzyme flexibility is required for catalytic function.

During experiments designed to determine the high-resolution crystal structure of a productive complex of ribonuclease A with a nucleotide substrate, we observed that the substrate failed to bind to the crystalline enzyme at about 215 K but seemed to react rapidly at temperatures only 20 K higher (B.F.R. *et al.*, data not shown). We presumed that this apparent loss of function was due to the change in dynamic behaviour observed in this and other proteins at around 200 K (Fig. 1). To test this hypothesis and to investigate the effect of the 200 K dynamical transition on enzyme function, we studied the crystallographic binding of a tight-binding competitive inhibitor just above and below the transition temperature. We used cytidine 2'-monophosphate (2'-CMP) because it binds to ribonuclease A with an association constant that is more than an order of magnitude tighter than that of the substrates⁷ and because it binds in the active site in a way that is analogous to substrate⁸.

A crystal of bovine pancreatic ribonuclease A was mounted in a flow cell⁹ and equilibrated with a cryoprotective mother liquor of 70% methanol/30% water at a proton activity of 5.5 (ref. 10). Ribonuclease is catalytically active in this mixed solvent down to at least 243 K, both in solution and in the crystalline state¹¹. The flow cell was mounted on a Siemens P3 single-counter four-circle diffractometer equipped with a modified LT-1 low-temperature device and a novel cryostat designed for long-term temperature stability during protein crystal X-ray diffraction data collection¹².

The crystal of ribonuclease A in the flow cell/cryostat system was cooled to 212 K (Fig. 1). Temperature stability was ± 1 K over a 3-week period. A set of diffraction data to 1.7 Å resolution was collected at this temperature, and the resulting electron density map examined in the active site region. Only the sulphate ion normally found to occupy the phosphate-binding subsite¹³ was observed (Fig. 2a). 2'-CMP at 2.2 mM, which corresponds to about 20 times its inhibition constant, K_i , was then passed over the crystal at a flow rate of 0.5 ml per day for six days at 212 K. The viscosity of the mother liquor is about 50 cP at this temperature¹⁰. During this period the intensities of reflections known to be sensitive to the binding of molecules in the specificity subsite were continuously monitored. No intensity changes were observed. After six days of exposure to inhibitor

at 212 K, a second set of data to 1.7 Å resolution was collected and another electron density map calculated. This map was virtually identical to the first (Fig. 2b), indicating that there had been no binding of 2'-CMP.

The temperature was raised to 228 K (Fig. 1) and the flow of 2'-CMP continued. At 228 K the viscosity of the mother liquor decreases to about 20 cP (ref. 10). Within minutes, intensity changes consistent with binding to the enzyme were observed; these changes were complete in less than two hours. A third set of data was collected at 1.7 Å resolution: the electron density map clearly showed the inhibitor molecule bound to the active site of the enzyme in the expected way (Fig. 2c).

To determine whether the inhibitor, once bound above the dynamical transition temperature, would fall off the enzyme as the transition was crossed, the temperature was dropped to 212 K and inhibitor-free mother liquor was passed over the crystal for two days. No intensity changes were observed. A fourth set of data to 1.7 Å resolution was collected, and this electron density map showed that the inhibitor remained bound (Fig. 2d). Finally, the temperature was raised to 228 K while the flow of inhibitor-free mother liquor was continued. Almost immediately, intensity changes characteristic of the loss of inhibitor from the active site were observed (data not shown).

At this point, the crystal had been maintained at subzero temperatures for almost a month and had been used to collect four high-resolution data sets. We checked to see whether anything could be bound to the protein below 220 K. Our observations could be explained by a phase transition in the mother liquor in the crystal interstices that prevents solute from reaching the enzyme below 220 K. To test this possibility, 10 mM K_2PtCl_4 was flowed over the crystal at 212 K. This compound reacts covalently at room temperature with the sulphur of methionine 29, which is on the surface of ribonuclease A (ref. 14). After overnight flow (~ 0.5 ml) at 212 K, a final set of data was collected, this time (owing to considerable cumulative radiation damage) only to 5 Å resolution. A difference electron density map between these data and calculated amplitudes from the native structure clearly showed platinum bound to Met 29 (Fig. 3). This result indicated that the dynamical transition at 220 K did not affect the reactivity of chemical groups on the protein surface. Only specific binding in the active-site pocket was abolished (or greatly reduced) below this temperature.

We conclude that below the dynamical transition at 220 K, crystalline ribonuclease A is unable to bind large, specific ligands in its active site. Binding may still be possible over longer periods

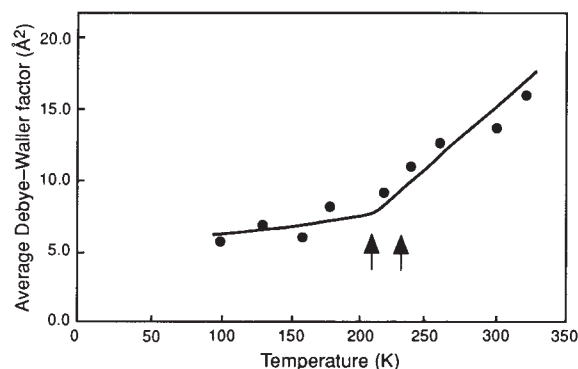
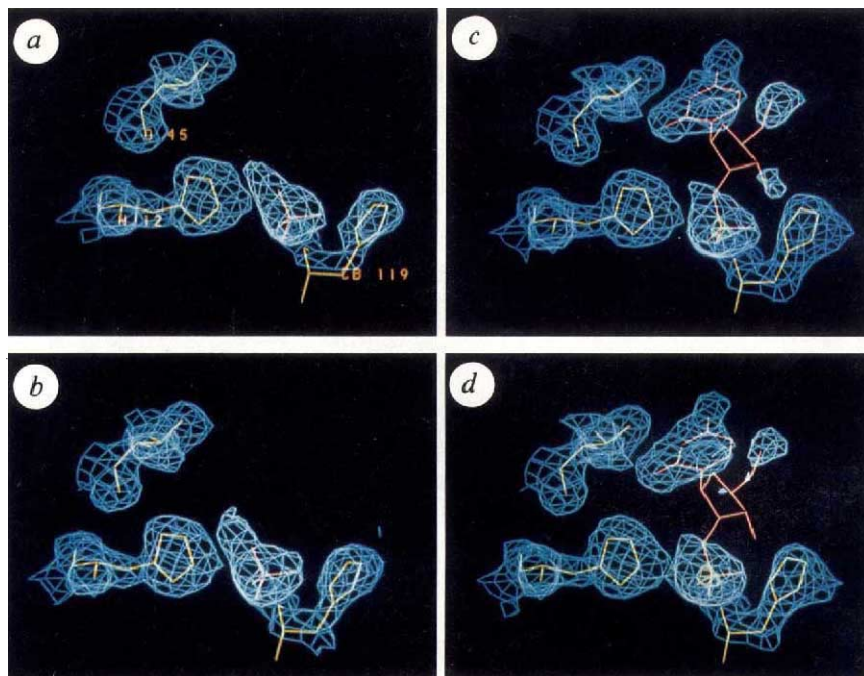


FIG. 1 Plot of the average Debye-Waller factor for all of the non-hydrogen protein atoms in crystalline bovine pancreatic ribonuclease A as a function of temperature (redrawn from data in ref. 1). The plot is biphasic, but the exact transition temperature is hard to estimate. This difficulty in part arises because the transition is broad, but there is also evidence that it depends on solvent composition and on the rate of cooling (see for example, ref. 16). Our estimate of the midpoint at 220 K is approximate; it is possible that the transition is still occurring at 200 K under the conditions of our experiment. We chose the two temperatures indicated by arrows for binding experiments so that they were as close as possible while still representing the two regions of behaviour.

FIG. 2 Electron density maps in the active site region of ribonuclease A. The maps are 'omit maps': the sidechains of the active-site residues and sulphate or inhibitor atoms were not included in the refinement or electron density calculations. Data were measured from a sorted list of intensities, with more counting time spent on the weaker reflections. Because it was necessary to measure so many data sets from one crystal, only the top 50% of the total unique data was collected. For each map the atomic coordinates of unliganded ribonuclease A at room temperature, with all active site groups omitted, were refined against the observed low-temperature X-ray data by the method of restrained least-squares. Refinement was terminated after about 10 cycles to avoid driving the structure into a false minimum. Stereochemistry was kept tight, with the r.m.s. deviation from ideal bond lengths being 0.014 Å and the r.m.s. deviation from ideal bond angles being 2°. *R* factors ranged from 0.22 to 0.20. No bound solvent molecules were included. After refinement, maps with coefficients $2F_{\text{obs}} - F_{\text{calc}}$ and calculated phases were examined. *a*, Electron density in the active site at 212 K before flowing in inhibitor. The two catalytic histidine residues, His 12 and His 119, are visible, as is Thr 45, the side-chain that hydrogen-bonds to the pyrimidine base⁷. Also seen is the sulphate ion that normally occupies the phosphate binding site¹³. Contours in this and subsequent maps are drawn at twice the estimated standard deviation. *b*, Electron density after 6 days' exposure to 2.2 mM 2'-CMP at 212 K. No inhibitor is found. *c*, Electron density after 2 h exposure to 2.2 mM 2'-CMP at 228 K. The inhibitor is observed in its expected C3'-endo conformation, identical to that found on the enzyme at room temperature⁸ but different from that of crystalline 2'-CMP alone¹⁷. The phosphate group occupies the sulphate



binding site and the base is hydrogen-bonded to the main-chain amide and side-chain hydroxyl of Thr 45. Little electron density is seen for the ribose at this contour level, because the sugar has almost no interaction with the enzyme and consequently is flexible. But density is present for all of the sugar atoms at a lower contour level. *d*, Electron density after 2 days' washing of the crystal with inhibitor-free mother liquor at 212 K. Very little reduction in electron density is seen.

than those investigated here, but in any case there is a large reduction in binding and dissociation rates over a very small range in temperature. Because the observed mean atomic positions for the protein atoms, especially in the active site, are identical within experimental precision in the 212 K and 228 K structures¹, we have attempted to explain these observations in terms of the dynamic behaviour of the enzyme. One explanation is that below 220 K the enzyme is too rigid to bind substrate or inhibitor rapidly. In the unbound structure of any enzyme, the residues in the active site do not all occupy exactly the positions they must have to allow substrate or inhibitor to fit into the pocket. Some adjustments are necessary, and apparently such motions are more complex, and have higher energy barriers, than simple harmonic vibrations. Below 220 K these collective

motions are 'frozen out' and the active site essentially consists of vibrating hard spheres, some of which are positioned to interfere with proper binding.

Another possibility is that below 220 K the water molecules bound in the active site (plus the sulphate ion in the case of ribonuclease) have become so rigid that they cannot be displaced easily by substrate or inhibitor. Still another explanation, which is almost a hybrid of the first two, is that below 220 K a phase transition occurs in the bound solvent shell around the protein¹⁵. This phase transition reduces the mobility of the bound water, and the change in solvent dynamics is then transmitted to the protein.

It may be feasible to generalize these results to other enzymes. Preliminary data indicate that crystalline porcine pancreatic elastase is also unable to bind a specific substrate below 210 K. Our results provide direct evidence that the thermal-driven collective atomic fluctuations in proteins are essential for rapid productive binding of large ligands. □

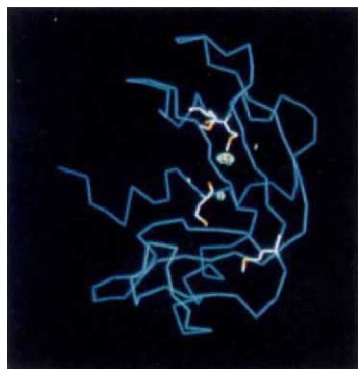


FIG. 3 Difference electron density map after exposure of the crystal to 10 mM potassium tetrachloroplatinate for 12 h at 212 K. The map was calculated with coefficients $F_{\text{obs}} - F_{\text{calc}}$, where the calculated *F*s and phases were based on the refined atomic coordinates for the native structure at 212 K. The map is contoured at 3 times the estimated standard deviation. Only the α -carbon backbone and methionine side chains are shown. The large spherical peak adjacent to the sulphur of Met 29 represents bound platinum.

Received 30 December 1991; accepted 28 April 1992.

1. Tilton, R. F. Jr, Dewan, J. C. & Petsko, G. A. *Biochemistry* **31**, 2469-2481 (1992).
2. Parak, F. et al. *FEBS Lett.* **117**, 368-372 (1980).
3. Keller, H. & Debrunner, P. *Phys. Rev. Lett.* **45**, 68-71 (1980).
4. Nocek, J. M. et al. *J. Am. chem. Soc.* **113**, 6822-6831 (1991).
5. Doster, W., Cusack, S. & Petry, W. *Nature* **337**, 754-756 (1989).
6. Loncharich, R. J. & Brooks, B. R. *J. molec. Biol.* **215**, 439-455 (1990).
7. Richards, F. M. & Wyckoff, H. W. *The Enzymes* 3rd edn Vol. IV (ed. Boyer, P. D.) 647-806 (1971).
8. Howlin, B., Harris, G. W., Moss, D. S. & Palmer, R. A. *J. molec. Biol.* **196**, 159-164 (1987).
9. Petsko, G. A. *Methods Enzym.* **114**, 141-146 (1985).
10. Douzou, P. *Cryobiology* (Academic, London, 1977).
11. Fink, A. L., Kar, D. & Kotin, R. *Biochemistry*, **26**, 8571-8579 (1987).
12. Rasmussen, B. F. & Petsko, G. A. *J. appl. Crystallogr.* (submitted).
13. Gilbert, W. A., Lord, R. C., Petsko, G. A. & Thammann, T. J. *J. Raman Spectrosc.* **12**, 173-179 (1982).
14. Carlisle, C. H., Palmer, R. A., Mazumdar, S. K., Gorinsky, B. A. & Yeates, D. G. R. *J. molec. Biol.* **85**, 1-18 (1974).
15. Doster, W., Bachleitner, A., Dunau, R., Hebl, M. & Lüscher, E. *Biophys. J.* **50**, 213-219 (1986).
16. Steinbach, P. J. et al. *Biochemistry* **30**, 3988-4001 (1991).
17. Kartha, G., Ambady, G. & Viswanamitra, M. A. *Science* **179**, 495-496 (1973).

ACKNOWLEDGEMENTS. We thank P. Douzou, A. Fink, H. Frauenfelder, M. Karplus, D. Koenig, S. Petersen, F. Richards and R. Tilton Jr for advice and encouragement. Supported by NIH grants (to G.A.P. and D.R.), by the Lucille P. Markey Charitable Trust (to G.A.P., D.R. and A.M.S.) and by NOVO Research Institute and the Carlsberg Laboratory (B.F.R.).